Cell Fate Decisions in Early Embryonic Development

Cell Fate Decisions in Early Embryonic Development

A dissertation presented

by

Xiaoxiao Zhang

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biochemistry

Harvard University

Cambridge, Massachusetts

April 2013

Dissertation Advisor: Andrew P. McMahon Xiaoxiao Zhang

Cell Fate Decisions in Early Embryonic Development

Abstract

The basis of developmental biology lies in the idea of when and how cells decide to divide or to differentiate. Previous studies have established several signaling pathways that determine cell fate decisions, including Notch, Wingless, Hedgehog, Bone morphogenetic protein, and Fibroblast growth factor. Signaling converges on transcriptional factors that regulate gene expression. In mouse embryonic stem cells, I explored how pluripotency and differentiation are regulated through opposing actions of β-catenin-mediated canonical Wnt signaling, and the mechanisms underlying Sonic hedgehog signaling in generating progenitor cells in the ventral neural tube.

In the first study, we uncovered a repertoire of β-catenin-associated *cis*-elements associated with the maintenance of pluripotency or early differentiation in mouse embryo stem cells. Which is preferred depends on transcription factors that partner with β-catenin. For pluripotency, β-catenin/Oct4/Tcf3 complex serves as an alternative to Oct4/Sox2 complex in maintaining self-renewal. For differentiation, β-catenin/Tcf/Lef complex acts through a more optimal Lef/Tcf motif to promote differentiation, which is blocked in the presence of an Erk inhibitor. These results point to a synergistic action of canonical Wnt signals with Fgf, or similar,

receptor tyrosine kinase dependent signals that utilize Erk-directed, Ets-mediated, transcriptional programs.

The second study focused on neural progenitors arising on differentiation of embryonic stem cells. We produced evidence supporting a central role for the general neural determinant Sox2 in establishing a neural-specific Sonic Hedgehog response. Their combined actions lead to the regional expression of distinct transcriptional regulators that determine ventral neural progenitor cell fates. Three of these are Nkx2.2, Nkx6.1 and Olig2: each has been shown to act as a repressor preventing premature differentiation and repressing alternative progenitor fates. A number of putative shared regulatory elements were identified, suggesting common sites of action for many of their functions. *In silico* study of Olig2 sequence specificity, in comparison with other basic-helix-loop-helix factors, leads to a model in which switching of Olig2 dimerization partners engages a distinct type of DNA binding motif.

In summary, this dissertation utilizes genome scale analyses to provide new mechanistic insights into cell signal regulated transcriptional programs that influence key cell fate decisions in early embryonic development.

Acknowledgement

I would like first to sincerely thank my advisor, Dr. Andrew McMahon, who provides me wonderful platform to work on exciting projects and support me throughout the past five years. Andy is an outstanding scientist, and is one of my most respected people. Andy has guided me to have a scientific way of thinking from experimental design to data interpretation to delivering presentation. I am proud of being part of the McMahon lab, and what I have learnt from Andy will continue to influence my life in the future. Thanks to my thesis committee, Drs. Catherine Dulac, Joshua Sanes, X. Shirley Liu, and Alex Meissner, for their helpful comments and critical inputs. Special thanks to Shirley, who kindly supervised me for over a year and provided me great opportunities to work with talented scientists and friends in her lab.

I would also like to thank all the members of the McMahon Lab and Liu Lab, who have discussed with me on projects and on life, with special thanks to Yuichi Nishi and Kevin Peterson, who patiently taught me experiments, as well as Shinsuke Ohba, from whom I learnt a lot on scientific rigorousness and excellence during our collaboration. Thanks also to Xinjun He, Ke Xu, Lori O'Brien, Jing Liu, Lick Lai, Bob Kao, Jingjing Guo, Jill McMahon, Joe Vanghan, and Celia Schneider from McMahon Lab, and Chongzhi Zang, Yiwen Chen, Su Wang, Sheng'en Hu, Cliff Meyer, Tao Liu, Wei Li, Han Xu, Hansen He, and Yin Liu from the Liu Lab. It has been a great pleasure working with such a wonderful group of people.

I am fortunate to have a group of very talented classmates at Harvard. All twenty of them are great people to interact with. Special thanks to Jessica Liu and Vu Nyugen, who have been supporting me in study and in life throughout the years. I wish them all the best for the future lives, and have no doubt they will excel in their fields. Special thanks to my dear friend Qingqing

Wang. During my most difficult times in Harvard, I was fortunate to have the friendship of her, who was always willing to lend me support, keep my spirits up, and share with me great moments.

I cannot thank enough for my family, who have always been alongside me, without whom I would certainly not be here. I thank my parents for their guidance throughout my life and especially for always cheering me up during my down time. Last but certainly not the least, I thank my husband, who has been accompanied me for my entire five years in the US. I thank him for his love, patience, dedication, and honesty. My family are the reason that supports me not to give up and I am excited to celebrate with them the end of this chapter of my life.

April 2013

Cambridge, MA

Table of Contents

List of Figures and Tables

List of Abbreviations

bHLH	Basic-loop-helix-loop
Bmp	Bone morphogenetic protein
ChIP-seq	Chromatin immunoprecipitation coupled with high throughput sequencing
CNS	Central nervous system
CRM	*cis*-regulatory module
EMT	Epithelial-mesenchymal transition
ERK	Extracellular signal-regulated kinases
ES (or ESC)	Embryonic stem (cell)
Fgf	Fibroblast growth factor
GBR	Gli1-binding region
GRN	Gene regulatory network
GSK	Glycogen synthase kinase
H3K4me1/2/3	Histone H3 lysine 4 mono-/di-/tri-methylation
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 tri-methylation
HD	Homeodomain
ICM	Inner cell mass
iPSC	induced pluripotent stem cell
LIF	Leukemia inhibitory factor
lincRNA	Large intergenic non-coding RNAs
MAPK	Mitogen-activated protein kinases
ME	Mesoderm
MEF	Mouse embryonic fibroblast
MN	Motor neuron
mRNA	Messenger RNA
NE	Neuroectoderm
NEB	Neural embryoid body
NT	Neural tube
OLP	Oligodendrocyte progenitor
PBM	Protein binding microarray
PCR	Polymerase chain reaction
PE	Primitive endoderm
PWM	Positional weight matrix
qPCR	Quantitative polymerase chain reaction
RT-qPCR	Real time quantitative polymerase chain reaction
Shh	Sonic Hedgehog
siRNA	Small interference RNA
SOM	Self-organizing map
TE	Trophectoderm
TF	Transcription factor
TGF- β	Transforming growth factor- β
VNT	Ventral neural tube
Wnt	Wingless

Chapter 1

Genomic analysis of gene regulatory networks underlying development

Abstract

The availability of whole-genome sequencing and advances in next-generation sequencing technology have transformed our approach to studying cell fate decisions during embryonic development. Gene regulatory networks have been elucidated by correlating genome-wide localization of transcription factors at *cis*-regulatory elements with gene expression patterns under different physiological conditions. Furthermore, extensive databases of epigenetic modifications and knowledge of crosstalk among signaling pathways add to our understanding of cell fate determination and cellular homeostasis. Together, these multi-pronged approaches allow for the characterization of embryonic development at the systems level in an unbiased manner.

Introduction

The response to different signaling pathways governs how a single cell divides and differentiates into a multi-cellular organism. A wide variety of cell types, each with distinct characteristics and functions, determine the ability of an organism to survive, thrive, and proliferate. Despite the morphological and functional variety, all cells in an organism share the same genomic material. Therefore, the regulatory processes governing gene expression determine distinct cellular difference such as when a cell replicates or differentiates into a more specialized cell type.

Our understanding of gene regulatory networks has accelerated with the advent of genome sequencing advances since the completion of the Human Genome Sequencing Project that have made direct sequence analysis at a genome scale a cost-effective option for the laboratory researcher (Lander et al., 2001; Venter et al., 2001) (Metzker, 2010; Shendure and Ji, 2008). Together with sophisticated computational and experimental strategies for genomic data processing (Burge and Karlin, 1997; Okazaki et al., 2002), a new era of genome-scale science has emerged. The application of next-generation sequencing (NGS) and advanced genomic, transcriptomic, proteomic, and epigenomic approaches have resulted in a tremendous amount of data (Hawkins et al., 2010). These genome-wide approaches eclipse studies at the single-gene level in identifying complete regulatory networks and identifying emergent properties that are only apparent at the systems level (Hasty et al., 2001; Mogilner et al., 2006).

The pluripotency networks that maintains embryonic stem cells (ESCs) serves as an excellent example. Several studies have mapped specific target genes of the a core pluripotency network made up of transcription factors (TFs) such as Nanog (Chambers et al., 2003; Mitsui et al., 2003), Oct4 (Nichols et al., 1998), and Sox2 (Avilion et al., 2003) using chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq). A comprehensive

understanding has emerged from these studies of a "pluripotency network" and signaling pathways that maintain this network (Boyer et al., 2005; Chen et al., 2008; Cole et al., 2008; Jiang et al., 2008; Kim et al., 2008; Liu et al., 2008; Loh et al., 2006; MacArthur et al., 2009; Muller et al., 2009). Interestingly, multiple regulatory factors bind at discrete regions of the genome and their action at these putative cis regulatory regions is generally presumed to govern local gene expression (Chen et al., 2008).

The regulatory actions have also been examined through other strategies. The Hi-C technique, a powerful tool for studying long-range DNA-DNA interactions, has characterized architectural dynamics of the human genome, which exists in complex open and closed chromatin states (Lieberman-Aiden et al., 2009). Mass spectroscopy using affinity purification has revealed protein-protein interaction networks including the core pluripotency factors in mESC (Ding et al., 2012; Gao et al., 2012; Liang et al., 2008; Mallanna et al., 2010; Pardo et al., 2010; van den Berg et al., 2010; Wang et al., 2006; Wu et al., 2006). Liver development has been shown to result from a complex GRN involving various hepatic TFs intertwined into a cross-regulatory and auto-regulatory network (Kyrmizi et al., 2006). Another intriguing and well-studied system is the intestinal epithelium, which exhibits an elegant balance of multi-potent intestinal stem cell self-renewal and differentiation (Sasai et al., 2012).

Multiple studies have uncovered signaling pathways involved in cell fate determination. Notch and Bmp signaling promotes differentiation of intestinal cell sub-types, while Hedgehog and Bmp signaling facilitates reciprocal communication between epithelial and mesenchymal cells (Crosnier et al., 2006; Madison et al., 2009). Wingless (Wnt) signaling functions in cell migration via Eph-ephrin signaling (Batlle et al., 2002); it also plays a role in proliferation of undifferentiated cells (Van Der Flier et al., 2007).

In this thesis, I studied the interplay of Wnt and Sonic Hedgehog (Shh) signaling in the determination of pluripotency in ESCs and the differentiation of ESCs to specific neural progenitors focusing on the GRNs triggered by these signaling pathways. The approach integrated our own data with the existing knowledge of *in vivo* protein-DNA and protein-protein interactions providing important insights into the GRNs governing maintenance and differentiation of stem cells, and the early programs of neural diversity.

In this opening chapter, I present an overview of early mouse embryonic development, as well as available genomic technologies that facilitate systematic analysis of underlying GRNs governing early cell states.

Biological system review

Early mouse embryonic development

A fertilized egg develops into the blastocyst over the course of 7 to 8 cell divisions (Figure 1.1A). When first formed, the blastocyst consists of an outer epithelium of trophectoderm (TE), whose descendants generate chorionic components of the placenta, and internally, a cluster of pluripotent cells called the inner cell mass (ICM). At implantation, the blastocyst is comprised of 128-256 cells. The ICM has progressed to pluripotent epiblast cells giving rise to an inner layer of primitive endoderm (PE) at the blastocoel surface. The PE ultimately generates additional extraembryonic cell types of the parietal and visceral yolk sacs while the epiblast gives rise to all the future somatic tissues of the adult mouse, the germ-line and additional cells for extraembryonic support functions. (Nowotschin and Hadjantonakis, 2010) (Figure 1.1A).

Figure 1.1

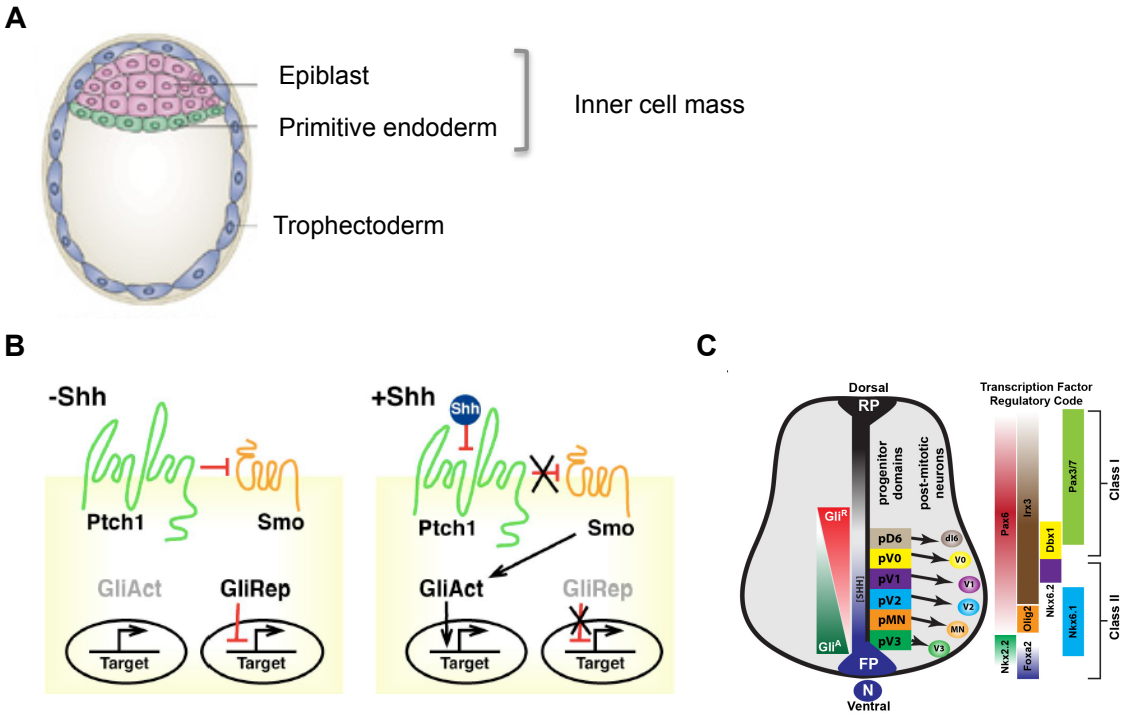

A
Epiblast
Primitive endoderm
Inner cell mass
Trophectoderm
B
-Shh
Ptch1
Smo
GliAct
GliRep
Target
Target
+Shh
Shh
Ptch1
Smo
GliAct
GliRep
Target
Target
C
Dorsal
RP
progenitor domains
post-mitotic neurons
GliR
GliA
[Shh]
pD6
pV0
pV1
pV2
pMN
pV3
dI6
V0
V1
V2
MN
V3
FP
N
Ventral
Transcription Factor Regulatory Code
Pax6
Irx3
Pax3/7
Dbx1
Nkx6.2
Olig2
Nkx2.2
Foxa2
Nkx6.1
Class I
Class II
D
t=0
t=24 hr
t=72 hr
Pre-induction
Shh+
Shh+
+RA
+SAG
-MEFs
+RA
Shh-
Shh-

Figure 1.1 (Continued)
***In vivo* early embryonic development and *in vitro* mESC differentiation into neural progenitor cells.**
(A) Schematic of a blastocyst stage embryo, highlighting PE, TE and epiblast. PE and epiblast together are derived from the earlier ICM. ESCs are a product of ICM culture.
(B) In the absence of ligand (-Shh), Ptch1 negatively regulates smoothened and the Gli repressor (Gli^{Rep}) form accumulates to silence target genes. In the presence of ligand (+Shh), Shh inhibits Ptch1 function allowing Gli TFs to become activated (Gli^{Act}). Gli activators translocate to the nucleus to positively regulate target genes. Importantly, this also results in the loss of Gli-mediated repression allowing genes to become activated by other non-Hedgehog dependent mechanisms.
(C) Schematic of a transverse section of an E10.5 mouse NT, and D-V progenitor domains established by the concentration gradient of various morphogens.
(D) Cell fate determinant TFs and expression domains are highlighted on the right. *In vitro* differentiation of mESCs to embryoid bodies (EBs) expressing different VNT markers under RA and SAG stimulation mimics the *in vivo* VNT cell fate specification.

Shortly after implantation, the proximal-distal and anterior-posteriors axes are established as a starting point for embryonic pattern formation. Gastrulation then starts around embryonic day 6. (E6.0). The appearance of the primitive streak is the first morphological signature (Robb and Tam, 2004; Stern et al., 1992; Tam and Behringer, 1997). Primitive streak cells form the progenitors for the three embryonic germ layers: endoderm, mesoderm, and ectoderm. Subsequently, various signaling pathways activate downstream effectors of different cell lineages. For example, nascent mesoderm is formed by a process called epithelial-mesenchymal transition (EMT) under Wnt, Bmp and Fgf signaling. Nodal, a ligand of the Transforming growth factor β (TGF-β) family, activates an endoderm-specific program demarcated by Sox17, FoxA2, and Hhex (Zorn and Wells, 2009). Finally, epiblast cells that fail to migrate through the primitive streak differentiate into ectodermal derivatives, including the surface layer of the skin and the central nervous system (Arnold and Robertson, 2009). Meanwhile, the germ-cell-specific program is initiated in pluripotent epiblast cells under control of inductive Bmp4 signaling, leading to formation of primordial germ cells (PGCs). At around E6.25 in the early post-implantation embryo, expression of the key factor Blimp1 marks the suppression of the somatic program and the activation of the PGC-specific germ-cell program (Lawson et al., 1999; Tremblay et al., 2001; Hayashi et al., 2002; Ohinata et al., 2005; Vincent et al., 2005).

Embryonic stem cells, pluripotency & transcription factors

All blastocyst lineages are capable of generating self-renewing stem cell lines in culture; namely, ESCs from the ICM, epiblast stem cells (EPI) from the epiblast, trophoblast stem cells (TSCs) from the TE, and extraembryonic endodermal cells (XEN-cells) from the visceral endoderm (Rossant, 2001). TSCs and XEN-cells can only differentiate into extraembryonic cells of their

lineages (Hemberger et al., 2004; Kunath et al., 2005; Stewart and Mintz, 1981; Tanaka et al., 1998). ESC and EPI cells, however, are pluripotent—they possess the capacity to give rise to all lineages in the living adult. Initially derived from mouse blastocyst in 1981 (Evans and Kaufman, 1981; Martin, 1981), and from the human blastocyst in 1998 (Thomson et al., 1998), ESCs have become a powerful tool for both scientific and clinical endeavors, especially in regenerative medicine and disease modeling.

ESCs have traditionally been cultured on a layer of 'feeder' cells in medium containing fetal calf serum to maintain pluripotency and stimulate cell division (Evans and Kaufman, 1981; Martin, 1981). Feeder cells were later found to produce a signal that inhibits ESC differentiation, which was identified as leukemia inhibitory factor (LIF) (Koopman and Cotton, 1984; Smith et al., 1988; Smith and Hooper, 1987; Smith and Hooper, 1983; Williams et al., 1988). LIF is a cytokine produced by the TE and normally required for implantation (Bhatt et al., 1991; Cheng et al., 2001; Stewart et al., 1992). LIF functions by binding a transmembrane receptor complex to activate Jak/Stat signaling, which results in Stat3 phosphorylation and ESC self-renewal (Matsuda et al., 1999; Niwa et al., 1998). Later experiments sought to replace the requirement for serum components. These studies identified bone morphogenetic factor-4 (BMP4), as a serum substitute for replication of ES cells with a 129 strain background (Ying et al., 2003). Most recently, the combination of a glycogen synthase kinase-3 (GSK3) inhibitor (CHIR99021) and Mitogen-activated protein kinases/Extracellular signal-regulated kinases (MAPK/ERK) inhibitors (PD0325901) have been shown to enable derivation and maintenance of mESCs in completely defined, feeder-free culture conditions. This is referred to as the '2i' culture system (Wray et al., 2010; Ying et al., 2008). More importantly, this system allows the derivation and culture of embryos of various backgrounds, including rat ESCs, which had previously defied

conventional derivation conditions (Li et al., 2008).

The various supplements and ingredients added to ESC cultures highlight distinct roles of specific signaling pathways in regulating pluripotency in ESCs. LIF acts through the STAT3 signaling pathway, which has been shown to maintain the undifferentiated state of ESCs by an active form of a (estrogen receptor) fusion protein (Matsuda et al., 1999), while Bmp4 acts via its downstream effector Smad1 to activate inhibitor of DNA binding (Id) regulatory factor family members, replacing serum requirements (Niwa et al., 1998; Ying et al., 2003). mESCs produce fibroblast growth factor 4 (Fgf4) and this is thought to stimulate differentiation via MEK/ERK signaling (Ma et al., 1992). Ets-family inhibition may underlie the action of PD03 in 2i medium (ref), though this remains to be demonstrated.

The role of Wnt signaling in development is significantly more complex (Sokol, 2011; Wray et al., 2011). This pathway acts during gastrulation, as revealed by analysis of Wnt3 and Wnt3a mutants (Liu et al., 1999; Ikeya et al., 2001), and β-catenin knockout mESCs and mouse strains (Haegel et al., 1995; Huelsken et al., 2000). β-catenin is the transcriptional activator and functions in a complex with Lef and Tcf in response to Wnt signaling. Moreover, by analyzing outgrowths from β-catenin knockout morulae, it was suggested that ESC formation and self-renewal did not require β-catenin, although β-catenin might be important for cell adhesion and forming compact cell colonies (Haegel et al., 1995; Huelsken et al., 2000). Thus early evidence suggested that canonical Wnt signaling is not required for ESC self-renewal.

However, increasing evidence is suggesting a positive role for Wnt/β-catenin pathway in the maintenance of mESC pluripotency. LIF, together with Wnt3a, an activator of β-catenin-directed canonical Wnt signaling, are reported to support ESC pluripotency in the absence of other factors (Ogawa et al., 2006a; Singla et al., 2006; Ten Berge et al., 2011). Involvement of β-

catenin is supported by the observation that its overexpression is sufficient to sustain self-renewal under low levels LIF (Ogawa et al., 2006b; Takao et al., 2007). Furthermore, CHIR-mediated stimulation of canonical Wnt signaling in the presence of PD03 in 2i medium blocks an intrinsic tendency of mouse ESCs to differentiate (Ying et al., 2008). In addition, BIO, another GSK-3 inhibitor, has been reported to maintain ESCs via up-regulation of LIF (Sato et al., 2004) and to enhance somatic cell-fusion-mediated somatic cell reprogramming through the accumulation of β-catenin (Lluis et al., 2008).

Wnt signaling also promotes reprogramming in induced pluripotent cells (iPSCs), substituting for c-Myc in the efficient propagation of iPSCs derived from mouse embryonic fibroblasts transfected with Sox2, Oct4, and Klf4 (Marson et al., 2008a). The down-regulation of "stemness marker genes" in ESCs lacking functional β-catenin supports a role of canonical Wnt signaling in maintenance of pluripotency (Anton et al., 2007).

However, the increased expression of Wnt/β-catenin target genes, including differentiation genes like T and Cdx1, in GSK3 double knockout mESCs also indicated the contribution of Wnt/β-catenin signaling to early lineage-specific differentiation (Doble et al., 2007). Moreover, during embryonic development, Wnt/β-catenin signaling is important for the formation of mesoderm from epiblast, the lack of which results in failure of EMT (Lindsley et al., 2006). Further, one study reported the stimulation of human ESC proliferation and differentiation by Wnt3a (Dravid et al., 2005). There is also substantial controversy on whether Tcf-dependent transcriptional activities from canonical Wnt signaling are required for Wnt pathway regulation of pluripotency. Lyashenko et al. (2011) highlighted β-catenin cell-adhesion function rather than signaling function as being significant in Wnt-mediated ESC maintenance (Lyashenko et al., 2011). Yi et al. (2011) and Wray et al. (2011) reported that the abrogation of Tcf3 repression, not

β-catenin-Tcf activation of target genes, is important for ESC self-renewal (Wray et al., 2011; Yi et al., 2011). Due to the conflicting views on the requirement of Wnt/β-catenin signaling in mESC maintenance versus embryonic development, its regulatory mechanism is of great interest.

An overarching question in the development field is how a unique pluripotency state is created and propagated compared with more differentiated cell types. At the end of various signaling pathways lie TFs, which are proteins that directly interact with DNA to regulate gene expression. The first known regulator of pluripotency, discovered about 15 years ago, is Pou5f1 (also known as Oct4), a member of the Pit-Oct-Unc (POU) family of homeodomain (HD) factors (Nichols et al., 1998). Knockout of Oct4 is pre-implantation lethal: TE forms but the ICM does not (Nichols et al., 1998). In mESCs, Oct4 knockout results in rapid differentiation, and interestingly, Oct4 dosage is tightly controlled. Oct4 depletion by 50% results in differentiation of mESCs into trophectodermal cells, while overexpression by 50% promotes mesodermal and endodermal differentiation (Niwa et al., 2000).

A later screen for factors capable of maintaining pluripotency in mESCs identified the HD protein Nanog (Chambers et al., 2007; Mitsui et al., 2003; Silva et al., 2009). While it is also expressed in the ICM, Nanog is suggested to be dispensable once pluripotency is achieved, as Nanog-null mESCs can retain pluripotency and do not commit to differentiation (Chambers et al., 2007; Mitsui et al., 2003; Silva et al., 2009). A third member of the core ESC pluripotency network, the high-mobility group box protein Sox2, shows a similar behavior to Oct4. Sox2 is expressed within the epiblast and the extra-embryonic ectoderm of the pre-implantation blastocyst. It is indispensable for the formation of a pluripotent ICM. Lack of Sox2 results in primarily trophectodermal cells (Avilion et al., 2003). The importance of Oct4 and Sox2 as core pluripotency factors is demonstrated through the generation of induced pluripotent stem cells

(iPSCs) from differentiated cells using Oct4, Sox2, c-Myc, and Klf4 (Maherali et al., 2007; Okita et al., 2007; Takahashi and Yamanaka, 2006; Wernig et al., 2007). Since then, combinations of TFs as well as small molecule substitutes have been developed to reprogram various cell types into iPSCs (Meissner et al., 2008; Nakagawa et al., 2008; Yu et al., 2007).

Besides the core pluripotency networks, additional pluripotency regulators have been discovered by various approaches, including analyses of protein-protein interactions (Ding et al., 2012; Gao et al., 2012; Liang et al., 2008; Mallanna et al., 2010; Pardo et al., 2010; van den Berg et al., 2010; Wang et al., 2006; Wu et al., 2006), large scale RNA interference-mediated gene knockdown (Ivanova et al., 2006), and genome-wide small interference (siRNA) screening (Ding et al., 2009). These factors include chromatin regulators, remodelers and modifiers, and DNA methyltransferases. Recently, MicroRNAs (miRNAs) and large intergenic non-coding RNAs (lincRNAs) have been shown to be involved in the ESC pluripotency network (Kanellopoulou et al., 2005; Wang et al., 2007b). For example, Marson et al. (2008) showed that Oct4, Sox2, Nanog, and Tcf3 bind at ESC-related miRNA targets and PcG-occupied tissue-specific miRNAs (Marson et al., 2008b). By loss-of-function studies on a number of ESC-expressed lincRNAs, Guttman et al. (2011) reported similar gene expression changes similar to the removal of ESC regulatory factors, suggesting key roles of lincRNAs in ESC self-renewal comparable to known core transcription factors (Guttman et al., 2011).

Morphogen and tissue patterning

Signals can act as both analog and digital information carriers. Morphogens are secreted signals that form a concentration gradient centered at the tissue-organizing center cells. This gradient allows for determination of more than one fate in response to levels of signal transduced over a

give period of time (Kicheva et al., 2012; Rogers and Schier, 2011). At increasing distances from the source, the signal gradually decreases, resulting in different cell fate determinants that are triggered by distinct signal concentrations. One example is the patterning of the Drosophila embryo along the anterior-posterior axis, which is set by the Bicoid TF that responds to the morphogen and controls downstream gap gene expression in a concentration-dependent manner (Stjohnston and Nussleinvolhard, 1992). In vertebrates, the formation of different progenitor cell types in the ventral neural tube (VNT) is the most widely studied example of how the Shh morphogen gradient activates different downstream genes at differential concentration threshold (Dessaud et al., 2008; Ingham and Placzek, 2006; Nishi et al., 2009; Ribes and Briscoe, 2009).

Early embryonic neural tube patterning and Shh signaling

The Hh signaling pathway has a broad developmental role through the regulation of cell proliferation, cell survival, and cell specification (Ingham and McMahon, 2001; Ingham and Placzek, 2006). Shh, Desert Hedgehog (Dhh), and Indian Hedgehog (Ihh) are the three Hh ligands identified in vertebrates (Echelard et al., 1993). Dhh is most closely related to Drosophila hedgehog, while Shh and Ihh are more similar to each other and also regulate ongoing tissue maintenance and repair in the adult organisms. Mis-regulation of these pathways can lead to diseases, such as cancer (Wang et al., 2007a).

As a processed signal peptide, the lipoprotein Shh contains a cholesterol group at the C-terminus and is palmitylated at the N-terminus. Within the VNT, Shh forms a concentration gradient emanating from two ventral sources, the notochord and floor plate (Grili Linde et al., 2001; Chamberlain et al., 2008). In the receiving cell, Hh ligands bind to the receptor Patched1 (Ptch1), a 12-pass transmembrane protein (Figure 1.1B). Ptch1 inhibits a 7-pass transmembrane

protein, Smoothened (Smo), an essential activator for all Hh targets (ref). This inhibition is relieved upon binding of the Hh ligand to Ptch1, allowing transduction of the signal across the membrane (Hooper and Scott, 2005; Ingham and McMahon, 2001; Jacob and Lum, 2007). The de-repressed Smo then activates Gli TFs, which are downstream effectors of Shh signaling. Gli TFs complete Shh signaling by promoting expression of target genes (Hooper and Scott, 2005).

In the vertebrate VNT, notochord-derived Shh induces different progenitor domains along the dorsal-ventral (D-V) axis. Here, a graded distribution of Shh is transformed into a gradient of Gli transcriptional activity that induces different neural progenitor domains (Dessaud et al., 2008; Stamataki et al., 2005) (Figure 1.1C). Each progenitor domain gives rise to distinct classes of interneurons or motor neurons (MNs). In the absence of Shh signal, the Gli-family TF Gli3 is processed to an N-terminal repressor form that mainly functions in the medial-to-ventral NT. In contrast, Shh input blocks the production of Gli repressor and generates Gli activator forms where signal levels are highest in the most ventral region of the NT (Ingham and McMahon, 2001). NT transcription factor determinants are classified into Class-I and Class-II, depending on their responses to Shh signaling. Class-I TFs are repressed by Shh, and Class-II TFs are activated by Shh. The combinatorial actions of Gli repressor and Gli activator are translated into transcriptional responses that activate Class-II transcriptional regulators and repress Class-I transcriptional regulators (Briscoe et al., 2000; Briscoe et al., 1999; Ericson et al., 1997; Persson et al., 2002). Class-II regulatory factors are fate determinants of ventral cell types. Class-I TFs mostly express in the dorsal NT, but a few of them are also expressed in the ventral NT.

More specifically, there are six distinct cell types in the VNT. From dorsal to ventral, these comprise five neural progenitor types (pV0, pV1, pV2, pMN and pV3) and a non-neuronal

floor plate (FP) (Ingham and McMahon, 2001) (Figure 1.1C). Neural progenitors will later give rise to different interneuron or MN cell types. Along the D-V axis of NT, each of the progenitor domain expresses different TFs. For example, from ventral to dorsal, Foxa2 species FP; Nkx2.2, pV3; Olig2, pMN; while Nkx6.1 is more broadly expressed from the FP to dorsal pV2 boundary. Ventral cell types, with the exception of the floor plate, emerge from dorsal to ventral, reflecting dynamics of the Shh signaling response where time and concentration are critical (Dessaud et al., 2007).

In addition to initial signal-mediated activation, reciprocal repression between TFs is thought to play a major role in defining sharp boundaries between adjacent progenitor domains (Briscoe et al., 2000; Briscoe et al., 1999; Ericson et al., 1997; Persson et al., 2002). For example, Shh signaling represses one group of dorsal TFs indirectly (Class I), while directly activates a second group within the VNT (Class II) (Briscoe et al., 2000). Cross-repression between the Class I TF Irx3 and the Class II TF Olig2 sharpens the boundary between immediately adjacent pV2 and pMN progenitor compartments. Similarly, the pV3 and pMN boundary is defined by cross-repression between the Class I Pax6 protein and Class II Nkx2.2 protein. In the simplest model, cross-repression is carried out directly by each regulatory factor. Most HD and basic-Helix-Loop-Helix (bHLH) proteins that act as ventral neural progenitor determinants are thought to act as repressors though the precise mechanisms of action have not been determined (Briscoe et al., 1999; Ericson et al., 1997; Mizuguchi et al., 2001; Muhr et al., 2001).

Besides Shh, other signaling molecules have also been reported to play a role in progenitor domain specification, such as Wnt (Ahn et al., 2008; Alvarez-Medina et al., 2009; Chesnutt et al., 2004; Lei et al., 2006; Megason and McMahon, 2002; Zechner et al., 2003), Bmp from the roof plate (Samanta and Kessler, 2004), retinoic acid (RA) (Novitch et al., 2003) and

Fgf signals from flanking somites (del Corral et al., 2003), and Notch signaling (Holmberg et al., 2008; Kageyama et al., 2008).

In vitro differentiation from mouse embryonic stem cells to neural progenitors

One advantage of utilizing an *in vitro* differentiation system is that mESCs can be used to provide an inexhaustible source of differentiated cell types, which is extremely important for genome-wide studies that require a substantial amount of cell material. However, it has been extremely challenging to directly differentiate ESCs into the desired lineage because of the complexity of the system. Recently, benefiting from growing understanding of various signaling pathways, ESCs have been shown to form specific germ-layer derivatives with high frequency (Fehling et al., 2003; Kubo, 2004; Ying et al., 2003). Methodologies for directed differentiation into specific cell-types have become increasingly sophisticated (Murry and Keller, 2008).

In the neural lineage, ESC have been shown to differentiate into NT derivatives by pharmaceutical intervention of culture conditions (Wichterle et al., 2002). In this study, mESCs were induced into neuronal lineage by RA induction, and were at the same time exposed to Shh pathway activation by a Smo agonist, SAG, a small molecule that acts as a potent activator of the pathway (Figure 1.1D) (Chen et al., 2002). This *in vitro* system has been shown to induce Class-II VNT genes in a time- and concentration-dependent manner, which mimics the *in vivo* patterning process (Wichterle et al., 2002). By adjusting SAG concentration, we are able to maximize the ventral cell types of interest, with a higher SAG concentration resulting in a more ventralized cell type.

Technology review

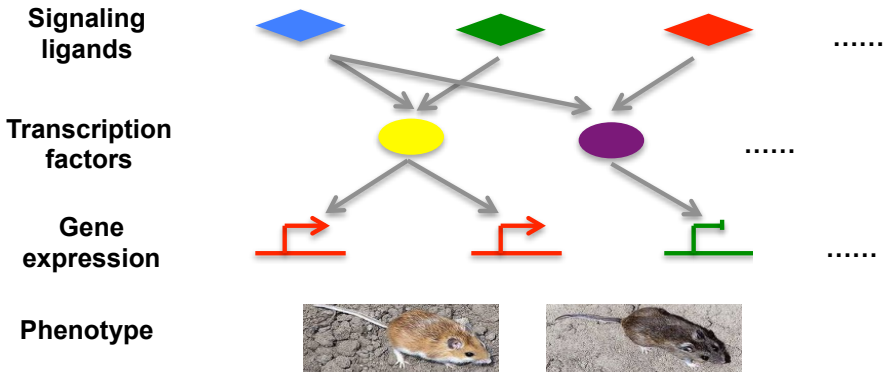
Genome-wide TF binding, epigenetic change & gene regulation

As the mechanistic foundation of cell fate decision, gene regulation has been the core interest of molecular biologists, the heart of which lies the study of TF target regulation (Figure 1.2A). TFs are the class of genes that control gene expression by direct binding at *cis*-regulatory elements via DNA binding domains (DBD) in a sequence-specific manner. Multiple pathways cross talk and converge at the level of TFs (Figure 1.2A) (Brivanlou and Darnell, 2002). Sequence-specific binding and antibody-mediated recognition of the target TF enables genome-wide immunoprecipitation of TF-associated DNA sequences. Analysis of these regions facilitates the prediction of sequence-specific binding feature of a TF. A metric called positional weight matrix (PWMs) defines a consensus DNA recognition motif, based on the DNA sequence enrichment in TF binding sites (Elnitski et al., 2006; Morgan et al., 2007; Wright and Funk, 1993).

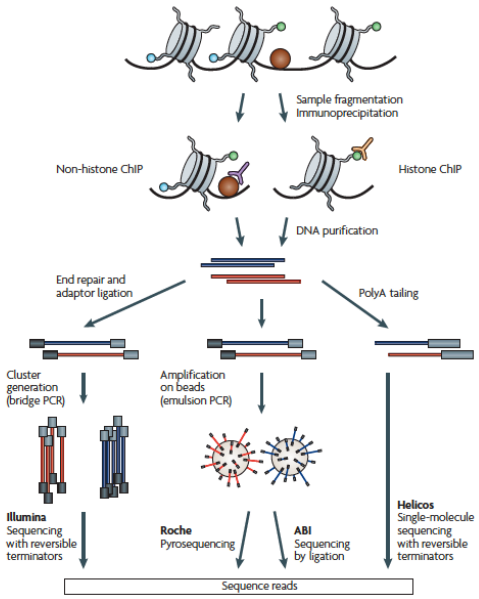
Most recently, local epigenetic profiles, defined as the chromatin states that are found at a given time point and cell type (Park, 2009; Wang et al., 2008), have been shown to be crucial in creating a permissive chromatin structure to allow accessibility of TFs. Different epigenetic modifications have been defined, but histone modifications, especially the combinatorial patterns of different histone modification codes, have been most widely studied. Histone octamers that allow the DNA to compact into nucleosomes are the major targets of various covalent modifications. For example, the N- and C-termini of histone H3 can be extensively methylated, phosphorylated, acetylated, and ubiquitinylated in a functionally important manner (Bernstein et al., 2007; Goldberg et al., 2010; Kouzarides, 2007; Lister et al., 2009). For example, methylation of histone 3 lysine 9 (H3K9), histone 3 lysine 27 (H3K27) and Histone 4 lysine 20 (H4K20) were first observed as playing a role in heterochromatin formation and gene repression

Figure 1.2

A



B



C

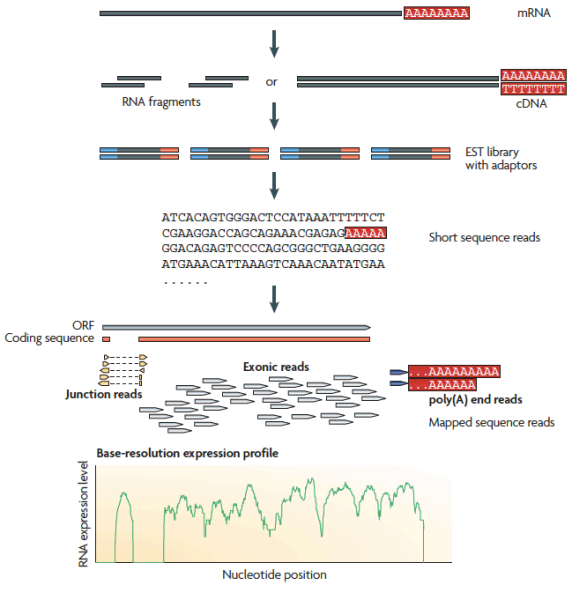


Figure 1.2 (Continued)
Using next-generation sequencing techniques to study gene regulatory networks.
(A) Schematic of developmental cues from signaling ligands to TF-mediated gene regulation to phenotypic changes as the read out.
(B) Standard ChIP-seq experiment. Adapted from Park et al. Figure 1 (Park, 2009).
(C) Standard RNA-seq experiment. Adapted from Wang et al. Figure 1 (Wang et al., 2009).

(Martens et al., 2005). One of the best characterized histone codes is the bivalent domains marked by histone 3 lysine 4 tri-methylation (H3K4me3) and histone 3 lysine 27 tri-methylation (H3K27me3), which marks key developmental genes that are repressed in ESCs but are poised for expression later in development (Bernstein et al., 2006). In addition, histone 3 lysine 27 acetylation (H3K27ac) has been shown by different studies to be an important active enhancer signature in both human and mouse, together with histone 3 lysine 4 mono-/di-methylation (H3K27me and H3K27me2) (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner et al., 2011).

By integrating global chromatin binding profiles for the same factor at different physiological conditions or of different TFs under the same condition, scientists are able to create a dynamic and integrated picture of TF-mediated GRNs in different developmental contexts. However, the information-rich nature of this idea is not achievable without the technology for sequencing DNA in a high-throughput manner.

ChIP-seq overview

Technological advances in NGS, defined as the use of established platforms such as Illumina/Solexa Genome Analyzer, Roche/454 Genome Sequencer and Applied Biosystems SOLiD platforms have made the collections of extensive DNA sequence information straight-forward and cost-effective (Metzker, 2010; Pareek et al., 2011; Shendure and Ji, 2008). Genome-wide mapping of protein-DNA interactions and epigenetic profiles have been studied extensively by ChIP-seq in the last five years (Figure 1.2B) (Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007). The procedure cross-links a protein, for example a TF or modified histone, to associated DNA. Antibodies against the TF or a histone

modification of interest are used to pull down the target proteins with locally cross-linked DNA. After a series of fragmentation and selection steps, DNA fragments approximately 200-600bp long are purified, cross-linking reversed, DNA fragments are size selected, ligated to common adaptors, and PCR amplified, then subjected to sequencing on one of the NGS platforms. The basic principles for NGS are similar: the adaptor-linked ChIP'ed DNA forms clonal clusters on the flow cell. During each step of enzyme-driven extension of the DNA template, fluorescent labels are incorporated and detected by high-resolution imaging. The output will be an unbiased consortium of sequences of various lengths ranging from 35bp to 200bp that the protein of interest interacts with (Shendure and Ji, 2008). Subsequent steps involve data processing and management. With various algorithms being produced and compared, each one bears its own advantage and limitations that should be chosen with caution depending on the desired purpose (Ji et al., 2008; Schones et al., 2008; Zhang et al., 2008).

DNA binding specificity and cell-type specific gene regulation

DNA recognition site specificity contains important information for *in vivo* TF functions. For example, using nucleosome resolution ChIP-seq of H3K4me1/2/3 under different androgen stimulations, He et al. (2010) predicted the involvement of Oct1 and Nkx3.1 in the androgen response of prostate cancer cells based on de novo motif discovery and validated the predictions experimentally (He et al., 2010). Within the same family of TFs, even minor differences in *in vitro* binding specificity may translate into *in vivo* selectivity. For example, Ets family factors all have similar DBDs (Wei et al., 2010). Ets factors can be categorized into four classes depending on DNA binding specificity, and the small differences in site recognition were validated *in vivo* by ChIP-seq. Similarly, despite the fact that NeuroD2 and MyoD are bHLH family proteins with

an E-box DNA recognition site, both can bind to different variants of E-box sites and regulate distinct lineage-specific gene regulation (Fong et al., 2012). Interestingly, the observation that TFs may recognize different sites in a context-dependent manner adds additional flexibility on top of DNA sequence recognition specificity. For example, one of the core pluripotency factors Oct4 was reported to bind to different motifs, depending on the presence or absence of Sox2 binding (Mason et al., 2010).

The study of sequence specificity benefits from the increase in genome-wide TF DNA specificity databases. Significant advances were made when a new DNA microarray-based technology called protein binding microarrays (PBMs) were developed, allowing for the sequence specificities of TFs to be characterized *in vitro* in a high-throughput manner (Badis et al., 2009; Berger and Bulyk, 2009; Bulyk, 2006). Briefly, the full length or DBD of the target TF is epitope-tagged, purified and applied to double-stranded DNA microarrays with a large number of DNA sequences designed in a way to exhaust all combinations of N-mer DNA sequences (Mukherjee et al., 2004). Then the protein-linked DNA microarray is labeled with a fluorophore-conjugated antibody against the epitope tag. After a series of washing, scanning, normalization and calculation, DNA sequences most significantly bound by the applied protein are then identified, which will be the final consensus or PWM for the *in vitro* binding specificity of the target protein. Compared to other techniques studying protein-DNA interactions, PBM has the advantage of being easily scalable and the array preparation and *in vitro* binding are straightforward.

Using this technique, Badis et al. (2009) examined 104 distinct mouse DNA binding proteins (Badis et al., 2009). The UniPROBE database, created from PBM technology and harboring DNA binding data for 406 non-redundant proteins ranging from a diverse range of

organisms has facilitated genome-wide studies for elucidating GRNs. Most recently, Jolma et al. (2013) conducted a comprehensive study on human TF sequence-specific binding using a high-throughput SELEX (HT-SELEX) (Jolma et al., 2010; Jolma et al., 2013), which is more useful for long motif studies in comparison with PBM. Their results add to our understanding of the DNA-protein binding specificity model, greatly expanded the TF DNA-binding database, and also revealed the importance of homodimer orientation, spacing preferences and base-stacking interactions on TF-DNA binding specificity.

As intriguing as it is, caution should be taken. First, the utilization of DBD motifs in studying TF binding an lies on the assumptions that the target sequences for a given TF is relatively invariant in different conditions or developmental stages. Second, this approach is not applicable to the scenario in which a detected binding event actually comes from chromatin looping from a more distal binding element. Third, the current notion of whether TFs bind directly or indirectly to chromatin, based on the presence or absence of TF binding motifs in peak regions. However, using PBMs, studies have shown that about half of the 104 mouse DNA-binding proteins recognize multiple different sequence motifs (Badis et al., 2009), suggesting that when analyzing motif data, researchers should be cautious when drawing conclusions that a protein of interest directly binds DNA. Finally, closely related family members are likely to recognize similar motif sequences, preventing the accurate identification of target TF simply based on motif binding information.

TF regulation other than sequence specificity

In the ideal situation, different TFs are expressed at different levels, and a characterization of the top expressed lineage-specific TF would serve as the hub and allow for prediction of a GRN for

a certain state. However, in reality, predictions and reconstruction of GRNs is complicated by factors such as chromatin accessibility, which is affected by various histone modifications. For example, a recent study highlighted the importance of chromatin accessibility in directing cell lineage differentiation determined by combinatorial histone modification codes (Fong et al., 2012). In this study, Neurod2 is a TF in neurons, while MyoD is a TF in skeletal muscle tissues. Neurod2 and MyoD direct differentiation to two completely distinct lineages, but the question is when both factors are expressed and how is a certain lineage chosen against the other. The authors show that epigenetic profiles in a particular cell state is determinant of which DNA sequences are exposed and then bound by one of the two factors. He et al. (2010) also demonstrated the predictive power of using only combinatorial histone modification signatures in predicting critical TFs for cell fate decisions (He et al., 2010).

Besides epigenomic influences, the analysis of GRNs is complicated by the observation that for some TFs, the number of genomic binding regions far exceed the known direct target genes. For example, one ChIP-seq study on MyoD uncovered 30000 to 60000 binding sites in skeletal muscle cells, most of which were not associated directly with target gene regulation (Cao et al., 2010). Instead, this study proposes a reciprocal effect of MyoD binding on local epigenomic changes. Thus besides the direct regulation of lineage specific gene expression, TFs can be involved in a broader diversity of activities that may not be directly associated with their target gene regulation; alternatively, TF can be involved in binding events without obvious transcriptional consequences.

RNA-seq overview

NGS sequencing has had a similar impact on our understanding of the transcriptome; the complete map RNA transcripts for a particular cell type in a certain physiological state; this includes mRNAs and non-coding RNAs in mammalian genome. The application of sequencing technology on transcriptome profiling, called RNA-seq (Figure 1.2C), is a huge advance, especially in identifying novel transcripts and in expanding our understanding of low abundance non-coding RNAs (Cloonan et al., 2008; Mortazavi et al., 2008). Briefly, this technique analyzes RNAs which are converted to complementary DNAs (cDNAs) and are ligated to adaptors at the ends. Similar to ChIP-seq, the samples are subjected to fragmentation either as RNA before reverse transcription or as cDNA. Short cDNA sequences of several hundred bps are then subjected to library construction and sequencing essentially as above. In comparison with the traditional hybridization-based quantification of the transcriptome, it is the first in which all transcripts, known or *de novo*, can be systematically detected and quantified without any prior knowledge. Associating TF binding to its target sites, and associating binding with functional significance, is facilitated by transcriptional profiling approaches including RNA-seq. However, connecting TF to target gene may not be straightforward when other co-factors are required besides the TF of interest, or when TF binding site needs to be directed to target promoter by mechanisms like looping.

Summary

The work presented in this thesis has addressed many foundational questions in the developmental biology field. Chapter 2 focuses on fate choice for mESCs. It describes using a genome-wide approach to investigate how β-catenin-mediated canonical Wnt signaling pathway

regulates mESCs self-renewal and early differentiation in a context dependent manner. The β-catenin-mediated GRN is tightly intertwined with core pluripotency network by Nanog/Oct4/Sox2. We propose a model where β-catenin contributes to pluripotency by utilizing *cis*-elements in common with core pluripotency factors. In addition, in the absence of pluripotency factors, β-catenin-mediated canonical Wnt signaling promotes differentiation gene expression together with the Fgf/ERK pathway effector Ets. In Chapter 3, cell fate decision-making is addressed in VNT progenitors in forming distinct neuronal progenitor cell types under the influence of the Shh pathway in collaboration with the pan-neural factor Sox2 through a genome-wide method. Further, neuronal subtype differentiation to oligodendrocyte progenitor cells from Olig2-expression progenitors is investigated, with a focus on different fate choices via sequence specific binding of different Olig2 dimer species. The observations and results from analyzing genomic data from these studies have added to our understanding of TF direct gene regulation via sequence specificity, TF chromatin binding in relation to genome accessibility, combinatorial regulation through multiple TFs, and the integration of expression and binding data to facilitate network studies. Chapter 4 summarizes the major conceptual advances that have stemmed from this thesis work and proposed follow-up directions based on current results in order to further mechanistic studies of inhibition of Fgf/ERK/Ets signaling in 2i-based mESC culture conditions. It will also allow for the exploration of RNA-seq data focused on Shh-responsive non-coding lincRNA in VNT patterning. This final chapter also provides a summary of challenges facing next-generation sequencing-based genomics and forward-looking statements on the future of GRN construction.

References

Ahn, S.M., Byun, K., Kim, D., Lee, K., Yoo, J.S., Kim, S.U., Jho, E.H., Simpson, R.J., and Lee, B. (2008). Olig2-Induced Neural Stem Cell Differentiation Involves Downregulation of Wnt Signaling and Induction of Dickkopf-1 Expression. Plos One *3*.

Alvarez-Medina, R., Le Dreau, G., Ros, M., and Marti, E. (2009). Hedgehog activation is required upstream of Wnt signalling to control neural progenitor proliferation. Development *136*, 3301-3309.

Anton, R., Kestler, H.A., and Kuhl, M. (2007). beta-Catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells. Febs Lett *581*, 5247-5254.

Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Bio *10*, 91-103.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Gene Dev *17*, 126-140.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X.*, et al.* (2009). Diversity and Complexity in DNA Recognition by Transcription Factors. Science *324*, 1720-1723.

Barski, A., Cuddapah, S., Cui, K.R., Roh, T.Y., Schones, D.E., Wang, Z.B., Wei, G., Chepelev, I., and Zhao, K.J. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Batlle, E., Henderson, J.T., Beghtel, H., van den Born, M.M.W., Sancho, E., Huls, G., Meeldijk, J., Robertson, J., van de Wetering, M., Pawson, T.*, et al.* (2002). Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell *111*, 251-263.

Berger, M.F., and Bulyk, M.L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat Protoc *4*, 393-411.

Bernstein, B., Meissner, A., and Lander, E. (2007). The Mammalian Epigenome. Cell *128*, 669-681.

Bernstein, B., Mikkelsen, T., Xie, X., Kamal, M., Huebert, D., Cuff, J., Fry, B., Meissner, A., Wernig, M., and Plath, K. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315-326.

Bhatt, H., Brunet, L.J., and Stewart, C.L. (1991). Uterine Expression of Leukemia Inhibitory Factor Coincides with the Onset of Blastocyst Implantation. P Natl Acad Sci USA *88*, 11408-11412.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.R., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Briscoe, J., Pierani, A., Jessell, T.M., and Ericson, J. (2000). A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. Cell *101*, 435-445.

Briscoe, J., Sussel, L., Serup, P., Hartigan-O'Connor, D., Jessell, T.M., Rubenstein, J.L., and Ericson, J. (1999). Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. Nature *398*, 622-627.

Brivanlou, A.H., and Darnell, J.E. (2002). Transcription - Signal transduction and the control of gene expression. Science *295*, 813-818.

Bulyk, M.L. (2006). DNA microarray technologies for measuring protein-DNA interactions. Curr Opin Biotech *17*, 422-430.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J Mol Biol *268*, 78-94.

Cao, Y., Yao, Z.Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., MacQuarrie, K.L., Davison, J., Morgan, M.T., Ruzzo, W.L.*, et al.* (2010). Genome-wide MyoD Binding in Skeletal Muscle Cells: A Potential for Broad Cellular Reprogramming. Dev Cell *18*, 662-674.

Chamberlain, C.E., Jeong, J., Guo, C., Allen, B.L., and McMahon, A.P. (2008). Notochord-derived Shh concentrates in close association with the apically positioned basal body in neural target cells and forms a dynamic gradient during neural patterning. Development *135*, 1097-1106.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-1234.

Chen, J.K., Taipale, J., Young, K.E., Maiti, T., and Beachy, P.A. (2002). Small molecule modulation of Smoothened activity. P Natl Acad Sci USA *99*, 14071-14076.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W.W., Jiang, J.M., *et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Cheng, J.G., Chen, J.R., Hernandez, L., Alvord, W.G., and Stewart, C.L. (2001). Dual control of LIF expression and LIF receptor function regulate Stat3 activation at the onset of uterine receptivity and embryo implantation. P Natl Acad Sci USA *98*, 8680-8685.

Chesnutt, C., Burrus, L.W., Brown, A.M.C., and Niswander, L. (2004). Coordinate regulation of neural tube patterning and proliferation by TGF beta and WNT activity. Dev Biol *274*, 334-347.

Cloonan, N., Forrest, A.R.R., Kolle, G., Gardiner, B.B.A., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., *et al.* (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nature Methods *5*, 613-619.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Gene Dev *22*, 746-755.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. P Natl Acad Sci USA *107*, 21931-21936.

Crosnier, C., Stamataki, D., and Lewis, J. (2006). Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. Nat Rev Genet *7*, 349-359.

del Corral, R.D., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., and Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. Neuron *40*, 65-79.

Dessaud, E., McMahon, A.P., and Briscoe, J. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. Development *135*, 2489-2503.

Ding, J.J., Guo, Y.B., Liu, S., Yan, Y.J., Chang, G., Kou, Z.H., Zhang, Y., Jiang, Y., He, F.C., Gao, S.R., *et al.* (2009). Embryonic stem cells derived from somatic cloned and fertilized blastocysts, are post-transcriptionally indistinguishable: A MicroRNA and protein profile comparison. Proteomics *9*, 2711-2721.

Ding, J.J., Xu, H.L., Faiola, F., Ma'ayan, A., and Wang, J.L. (2012). Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res *22*, 155-167.

Doble, B.W., Patel, S., Wood, G.A., Kockeritz, L.K., and Woodgett, J.R. (2007). Functional Redundancy of GSK-3α and GSK-3β in Wnt/β-Catenin Signaling Shown by Using an Allelic Series of Embryonic Stem Cell Lines. Dev Cell *12*, 957-971.

Dravid, G., Ye, Z., Hammond, H., Chen, G., Pyle, A., Donovan, P., Yu, X., and Cheng, L. (2005). Defining the Role of Wnt/ β -Catenin Signaling in the Survival, Proliferation, and Self-Renewal of Human Embryonic Stem Cells. Stem Cells *23*, 1489-1501.

Echelard, Y., Epstein, D.J., Stjacques, B., Shen, L., Mohler, J., Mcmahon, J.A., and Mcmahon, A.P. (1993). Sonic-Hedgehog, a Member of a Family of Putative Signaling Molecules, Is Implicated in the Regulation of Cns Polarity. Cell *75*, 1417-1430.

Elnitski, L., Jin, V.X., Farnham, P.J., and Jones, S.J.M. (2006). Locating mammalian transcription factor binding sites: A survey of computational and experimental techniques. Genome Research *16*, 1455-1464.

Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., van Heyningen, V., Jessell, T.M., and Briscoe, J. (1997). Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. Cell *90*, 169-180.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in Culture of Pluripotential Cells from Mouse Embryos. Nature *292*, 154-156.

Fehling, H.J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G., and Kouskoff, V. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. Development *130*, 4217-4227.

Fong, A.P., Yao, Z.Z., Zhong, J.W., Cao, Y., Ruzzo, W.L., Gentleman, R.C., and Tapscott, S.J. (2012). Genetic and Epigenetic Determinants of Neurogenesis and Myogenesis. Dev Cell *22*, 721-735.

Gao, Z.H., Cox, J.L., Gilmore, J.M., Ormsbee, B.D., Mallanna, S.K., Washburn, M.P., and Rizzino, A. (2012). Determination of Protein Interactome of Transcription Factor Sox2 in Embryonic Stem Cells Engineered for Inducible Expression of Four Reprogramming Factors. J Biol Chem *287*, 11384-11397.

Goldberg, A.D., Banaszynski, L.A., Noh, K.-M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X*., et al.* (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. Cell *140*, 678-691.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L*., et al.* (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295-U260.

Haegel, H., Larue, L., Ohsugi, M., Fedorov, L., Herrenknecht, K., and Kemler, R. (1995). Lack of beta-catenin affects mouse development at gastrulation. Development *121*, 3529-3537.

Hasty, J., Isaacs, F., Dolnik, M., McMillen, D., and Collins, J.J. (2001). Designer gene networks: Towards fundamental cellular control. Chaos *11*, 207-220.

Hawkins, R.D., Hon, G.C., and Ren, B. (2010). Next-generation genomics: an integrative approach. Nat Rev Genet *11*, 476-486.

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M.*, et al.* (2010). Nucleosome dynamics define transcriptional enhancers. Nat Genet *42*, 343-347.

Hemberger, M., Hughes, M., and Cross, J.C. (2004). Trophoblast stem cells differentiate in vitro into invasive trophoblast giant cells. Dev Biol *271*, 362-371.

Holmberg, J., Hansson, E., Malewicz, M., Sandberg, M., Perlmann, T., Lendahl, U., and Muhr, J. (2008). SoxB1 transcription factors and Notch signaling use distinct mechanisms to regulate proneural gene function and neural progenitor differentiation. Development *135*, 1843-1851.

Hooper, J.E., and Scott, M.P. (2005). Communicating with Hedgehogs. Nat Rev Mol Cell Bio *6*, 306-317.

Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C., and Birchmeier, W. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. J Cell Biol *148*, 567-578.

Ingham, P.W., and McMahon, A.P. (2001). Hedgehog signaling in animal development: paradigms and principles. Gene Dev *15*, 3059-3087.

Ingham, P.W., and Placzek, M. (2006). Orchestrating ontogenesis: variations on a theme by sonic hedgehog. Nat Rev Genet *7*, 841-850.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., Decoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533-538.

Jacob, L., and Lum, L. (2007). Deconstructing the hedgehog pathway in development and disease. Science *318*, 66-68.

Ji, H.K., Jiang, H., Ma, W.X., Johnson, D.S., Myers, R.M., and Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat Biotechnol *26*, 1293-1300.

Jiang, J.M., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353-U103.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science *316*, 1497-1502.

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G.H., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpaa, M.J.*, et al.* (2010). Multiplexed massively parallel SELEX for

characterization of human transcription factor binding specificities. Genome Research *20*, 861-873.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G.H., *et al.* (2013). DNA-Binding Specificities of Human Transcription Factors. Cell *152*, 327-339.

Kageyama, R., Ohtsuka, T., Shimojo, H., and Imayoshi, I. (2008). Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition. Nat Neurosci *11*, 1247-1251.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Gene Dev *19*, 489-501.

Kicheva, A., Cohen, M., and Briscoe, J. (2012). Developmental Pattern Formation: Insights from Physics and Biology. Science *338*, 210-212.

Kim, S., Lee, S.G., Moon, S.H., Kim, J., Lee, S.H., and Chung, H.M. (2008). Developmental pattern of human embryonic stem cells in optimized hypoxic culture condition. Tissue Eng Regen Med *5*, 474-481.

Koopman, P., and Cotton, R.G.H. (1984). A Factor Produced by Feeder Cells Which Inhibits Embryonal Carcinoma Cell-Differentiation - Characterization and Partial-Purification. Exp Cell Res *154*, 233-242.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kubo, A. (2004). Development of definitive endoderm from embryonic stem cells in culture. Development *131*, 1651-1662.

Kunath, T., Arnaud, D., Uy, G.D., Okamoto, L., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R.L., Avner, P., and Rossant, J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development *132*, 1649-1661.

Kyrmizi, I., Hatzis, P., Katrakili, N., Tronche, F., Gonzalez, F.J., and Talianidis, I. (2006). Plasticity and expanding complexity of the hepatic transcription factor network during liver development. Gene Dev *20*, 2293-2305.

Lander, E.S., Consortium, I.H.G.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lei, Q.B., Jeong, Y., Misra, K., Li, S.K., Zelman, A.K., Epstein, D.J., and Matise, M.P. (2006). Wnt signaling inhibitors regulate the transcriptional response to morphogenetic Shh-Gli signaling in the neural tube. Dev Cell *11*, 325-337.

Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.L.*, et al.* (2008). Germline competent embryonic stem cells derived from rat blastocysts. Cell *135*, 1299-1310.

Liang, J., Wan, M., Zhang, Y., Gu, P.L., Xin, H.W., Jung, S.Y., Qin, J., Wong, J.M., Cooney, A.J., Liu, D.*, et al.* (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol *10*, 731-739.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science *326*, 289-293.

Lindsley, R.C., Gill, J.G., Kyba, M., Murphy, T.L., and Murphy, K.M. (2006). Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. Development *133*, 3787-3796.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Liu, X.S., Huang, J.Y., Chen, T., Wang, Y., Xin, S.M., Li, J., Pei, G., and Kang, J.H. (2008). Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells. Cell Res *18*, 1177-1189.

Lluis, F., Pedone, E., Pepe, S., and Cosma, M.P. (2008). Periodic Activation of Wnt/beta-Catenin Signaling Enhances Somatic Cell Reprogramming Mediated by Cell Fusion. Cell Stem Cell *3*, 493-507.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W.W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J.*, et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

Lyashenko, N., Winter, M., Migliorini, D., Biechele, T., Moon, R.T., and Hartmann, C. (2011). Differential requirement for the dual functions of β-catenin in embryonic stem cell self-renewal and germ layer formation. Nat Cell Biol *13*, 753-761.

Ma, Y.G., Rosfjord, E., Huebert, C., Wilder, P., Tiesman, J., Kelly, D., and Rizzino, A. (1992). Transcriptional Regulation of the Murine K-Fgf Gene in Embryonic-Cell Lines. Dev Biol *154*, 45-54.

MacArthur, B.D., Ma'ayan, A., and Lemischka, I.R. (2009). Systems biology of stem cell fate and cellular reprogramming. Nat Rev Mol Cell Bio *10*, 672-681.

Madison, B.B., McKenna, L.B., Dolson, D., Epstein, D.J., and Kaestner, K.H. (2009). FoxF1 and FoxL1 Link Hedgehog Signaling and the Control of Epithelial Proliferation in the Developing Stomach and Intestine. J Biol Chem *284*, 5936-5944.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., and Jaenisch, R. (2007). Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. Cell Stem Cell *1*, 55-70.

Mallanna, S.K., Ormsbee, B.D., Iacovino, M., Gilmore, J.M., Cox, J.L., Kyba, M., Washburn, M.P., and Rizzino, A. (2010). Proteomic Analysis of Sox2-Associated Proteins During Early Stages of Mouse Embryonic Stem Cell Differentiation Identifies Sox21 as a Novel Regulator of Stem Cell Fate. Stem Cells *28*, 1715-1727.

Marson, A., Foreman, R., Chevalier, B., Bilodeau, S., Kahn, M., Young, R.A., and Jaenisch, R. (2008a). Wnt signaling promotes reprogramming of somatic cells to pluripotency. Cell Stem Cell *3*, 132-135.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008b). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Martens, J.H.A., O'sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. Embo J *24*, 800-812.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci USA *78*, 7634-7638.

Mason, M.J., Plath, K., and Zhou, Q. (2010). Identification of Context-Dependent Motifs by Contrasting ChIP Binding Data. Bioinformatics *26*, 2826-2832.

Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. Embo J *18*, 4261-4269.

Megason, S.G., and McMahon, A.P. (2002). A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. Development *129*, 2087-2098.

Meissner, A., Mikkelsen, T.S., Gu, H.C., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X.L., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-U791.

Metzker, M.L. (2010). Applications of Next-Generation Sequencing Sequencing Technologies - the Next Generation. Nat Rev Genet *11*, 31-46.

Mikkelsen, T.S., Ku, M.C., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-U552.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Mizuguchi, R., Sugimori, M., Takebayashi, H., Kosako, H., Nagao, M., Yoshida, S., Nabeshima, Y., Shimamura, K., and Nakafuku, M. (2001). Combinatorial roles of olig2 and neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. Neuron *31*, 757-771.

Mogilner, A., Wollman, R., and Marshall, W.F. (2006). Quantitative modeling in cell biology: What is it good for? Dev Cell *11*, 279-287.

Morgan, X.C., Ni, S., Miranker, D.P., and Iyer, V.R. (2007). Predicting combinatorial binding of transcription factors to regulatory elements in the human genome by association rule mining. Bmc Bioinformatics *8*.

Mortazavi, A., Williams, B.A., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Meth *5*, 621-628.

Muhr, J., Andersson, E., Persson, M., Jessell, T.M., and Ericson, J. (2001). Groucho-mediated transcriptional repression establishes progenitor cell pattern and neuronal fate in the ventral neural tube. Cell *104*, 861-873.

Mukherjee, S., Berger, M.F., Jona, G., Wang, X.S., Muzzey, D., Snyder, M., Young, R.A., and Bulyk, M.L. (2004). Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nat Genet *36*, 1331-1339.

Muller, T., Fleischmann, G., Eildermann, K., Matz-Rensing, K., Horn, P.A., Sasaki, E., and Behr, R. (2009). A novel embryonic stem cell line derived from the common marmoset monkey (Callithrix jacchus) exhibiting germ cell-like characteristics. Hum Reprod *24*, 1359-1372.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: Lessons from embryonic development. Cell *132*, 661-680.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol *26*, 101-106.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.

Nishi, Y., Ji, H., Wong, W.H., McMahon, A.P., and Vokes, S.A. (2009). Modeling the spatio-temporal network that drives patterning in the vertebrate central nervous system. Biochim Biophys Acta *1789*, 299-305.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Gene Dev *12*, 2048-2060.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

Novitch, B.G., Wichterle, H., Jessell, T.M., and Sockanathan, S. (2003). A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. Neuron *40*, 81-95.

Nowotschin, S., and Hadjantonakis, A.K. (2010). Cellular dynamics in the early mouse embryo: from axis formation to gastrulation. Curr Opin Genet Dev *20*, 420-427.

Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D., and Niwa, H. (2006a). Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. Biochem Bioph Res Co *343*, 159-166.

Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D., and Niwa, H. (2006b). Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. Biochem Biophys Res Commun *343*, 159-166.

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H.*, et al.* (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature *420*, 563-573.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. Cell Stem Cell *6*, 382-395.

Pareek, C.S., Smoczynski, R., and Tretyn, A. (2011). Sequencing technologies and genome sequencing. J Appl Genet *52*, 413-435.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet *10*, 669-680.

Persson, M., Stamataki, D., te Welscher, P., Andersson, E., Böse, J., Rüther, U., Ericson, J., and Briscoe, J. (2002). Dorsal-ventral patterning of the spinal cord requires Gli3 transcriptional repressor activity. Gene Dev *16*, 2865-2878.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Ribes, V., and Briscoe, J. (2009). Establishing and Interpreting Graded Sonic Hedgehog Signaling during Vertebrate Neural Tube Patterning: The Role of Negative Feedback. Csh Perspect Biol *1*.

Robb, L., and Tam, P.P.L. (2004). Gastrula organiser and embryonic patterning in the mouse. Seminars in Cell & Developmental Biology *15*, 543-554.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A.*, et al.* (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods *4*, 651-657.

Rogers, K.W., and Schier, A.F. (2011). Morphogen Gradients: From Generation to Interpretation. Annual Review of Cell and Developmental Biology, Vol 27 *27*, 377-407.

Rossant, J. (2001). Stem cells from the mammalian blastocyst. Stem Cells *19*, 477-482.

Samanta, J., and Kessler, J.A. (2004). Interactions between ID and OLIG proteins mediate the inhibitory effects of BMP4 on oligodendroglial differentiation. Development *131*, 4131-4142.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med *10*, 55-63.

Schones, D., Cui, K., Cuddapah, S., Roh, T., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell *132*, 887-898.

Shendure, J., and Ji, H.L. (2008). Next-generation DNA sequencing. Nat Biotechnol *26*, 1135-1145.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., Oosten, A.L.v., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. Cell *138*, 722-737.

Singla, D.K., Schneider, D.J., LeWinter, M.M., and Sobel, B.E. (2006). wnt3a but not wnt11 supports self-renewal of embryonic stem cells. Biochem Bioph Res Co *345*, 789-795.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of Pluripotential Embryonic Stem-Cell Differentiation by Purified Polypeptides. Nature *336*, 688-690.

Smith, A.G., and Hooper, M.L. (1987). Buffalo Rat-Liver Cells Produce a Diffusible Activity Which Inhibits the Differentiation of Murine Embryonal Carcinoma and Embryonic Stem-Cells. Dev Biol *121*, 1-9.

Smith, T.A., and Hooper, M.L. (1983). Medium Conditioned by Feeder Cells Inhibits the Differentiation of Embryonal Carcinoma Cultures. Exp Cell Res *145*, 458-461.

Sokol, S.Y. (2011). Maintaining embryonic stem cell pluripotency with Wnt signaling. Development *138*, 4341-4350.

Stamataki, D., Ulloa, F., Tsoni, S.V., Mynett, A., and Briscoe, J. (2005). A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. Gene Dev *19*, 626-641.

Stern, C.D., Hatada, Y., Selleck, M.A.J., and Storey, K.G. (1992). Relationships between Mesoderm Induction and the Embryonic Axes in Chick and Frog Embryos. Development, 151-156.

Stewart, C.L., Kaspar, P., Brunet, L.J., Bhatt, H., Gadi, I., Kontgen, F., and Abbondanzo, S.J. (1992). Blastocyst Implantation Depends on Maternal Expression of Leukemia Inhibitory Factor. Nature *359*, 76-79.

Stewart, T.A., and Mintz, B. (1981). Successive Generations of Mice Produced from an Established Culture Line of Euploid Teratocarcinoma Cells. P Natl Acad Sci-Biol *78*, 6314-6318.

Stjohnston, D., and Nussleinvolhard, C. (1992). The Origin of Pattern and Polarity in the Drosophila Embryo. Cell *68*, 201-219.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663-676.

Takao, Y., Yokota, T., and Koide, H. (2007). Beta-catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. Biochem Biophys Res Commun *353*, 699-705.

Tam, P.P.L., and Behringer, R.R. (1997). Mouse gastrulation: the formation of a mammalian body plan. Mech Develop *68*, 3-25.

Tanaka, S., Takahashi, T., Takayanagi, H., Miyazaki, T., Oda, H., Nakamura, K., Hirai, H., and Kurokawa, T. (1998). Modulation of osteoclast function by adenovirus vector induced epidermal growth factor receptor. J Bone Miner Res *13*, 1714-1720.

Ten Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., Siu, R.K., and Nusse, R. (2011). Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. Nat Cell Biol *13*, 1070-U1088.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. Cell Stem Cell *6*, 369-381.

Van Der Flier, L.G., Sabates-Bellver, J., Oving, I., Haegebarth, A., De Palo, M., Anti, M., Van Gijn, M.E., Suijkerbuijk, S., Van De Wetering, M., Marra, G.*, et al.* (2007). The intestinal Wnt/TCF signature. Gastroenterology *132*, 628-632.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-+.

Wang, J.L., Rao, S., Chu, J.L., Shen, X.H., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Wang, Y., McMahon, A.P., and Allen, B.L. (2007a). Shifting paradigms in Hedgehog signaling. Curr Opin Cell Biol *19*, 159-165.

Wang, Y.M., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007b). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet *39*, 380-385.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.

Wang, Z.B., Zang, C.Z., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K.R., Roh, T.Y., Peng, W.Q., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.

Wei, G.-H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R.*, et al.* (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. Embo J *29*, 2147-2160.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-324.

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385-397.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid-Leukemia Inhibitory Factor Maintains the Developmental Potential of Embryonic Stem-Cells. Nature *336*, 684-687.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency

network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol *13*, 838-U246.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol *13*, 838-845.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem Soc Trans *38*, 1027-1032.

Wright, W.E., and Funk, W.D. (1993). Casting for Multicomponent DNA-Binding Complexes. Trends in Biochemical Sciences *18*, 77-80.

Wu, S., Wu, Y., and Capecchi, M.R. (2006). Motoneurons and oligodendrocytes are sequentially generated from neural stem cells but do not appear to share common lineage-restricted progenitors in vivo. Development *133*, 581-590.

Yi, F., Pereira, L., Hoffman, J.A., Shy, B.R., Yuen, C.M., Liu, D.R., and Merrill, B.J. (2011). Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nat Cell Biol *13*, 762-770.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-U515.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, et al.* (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science *318*, 1917-1920.

Zechner, D., Fujita, Y., Hulsken, J., Muller, T., Walther, I., Taketo, M.M., Crenshaw, E.B., Birchmeier, W., and Birchmeier, C. (2003). beta-catenin signals regulate cell growth and the balance between progenitor cell expansion and differentiation in the nervous system. Dev Biol *258*, 406-418.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Research *21*, 1273-1283.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol *9*.

Zorn, A.M., and Wells, J.M. (2009). Vertebrate Endoderm Development and Organ Formation. Annu Rev Cell Dev Bi *25*, 221-251.

Chapter 2

Gene regulatory networks mediating canonical Wnt signal directed control of pluripotency and differentiation in embryo stem cells

Addendum

This chapter was originally published in Stem Cells as:

Xiaoxiao Zhang*, Kevin A. Peterson, X. Shirley Liu, Andrew P. McMahon, & Shinsuke Ohba* (2013). Gene regulatory networks mediating canonical Wnt signal directed control of pluripotency and differentiation in embryo stem cells. Stem Cells. 2013 Mar 15. doi: 10.1002/stem.1371.

*These authors contributed equally to this work.

Author contributions: X.Z. and S.O.: conception and design, collection and/or assembly of data, data analysis and interpretation, manuscript writing; K.A.P.: collection and/or assembly of data, data analysis and interpretation; X.S.L.: data analysis and interpretation; APM: conception, design and interpretation, financial support, manuscript writing, final approval of manuscript.

Abstract

Canonical Wnt signaling supports the pluripotency of embryonic stem cells (ESCs) but also promotes differentiation of early mammalian cell lineages. To explain these paradoxical observations, we explored the gene regulatory networks at play. Canonical Wnt signaling is intertwined with the pluripotency network comprising Nanog, Oct4, and Sox2 in mouse ESCs. In defined media supporting the derivation and propagation of ESCs, Tcf3 and β-catenin interact with Oct4; Tcf3 binds to Sox motif within Oct-Sox composite motifs that are also bound by Oct4-Sox2 complexes. Further, canonical Wnt signaling up-regulates the activity of the *Pou5f1* distal enhancer via the Sox motif *in ESCs*. When viewed in the context of published studies on Tcf3 and β-catenin mutants, our findings suggest Tcf3 counters pluripotency by competition with Sox2 at these sites, and Tcf3 inhibition is blocked by β-catenin entry into this complex. Wnt pathway stimulation also triggers β-catenin association at regulatory elements with classic Lef/Tcf motifs associated with differentiation programs. The failure to activate these targets in the presence of a MEK/ERK inhibitor essential for ESC culture suggests MEK/ERK signaling and canonical Wnt signaling combine to promote ESC differentiation.

Introduction

A central question in all stem cell-based systems is how the balance of stem cell maintenance and commitment is regulated. ESCs derived directly from the early mammalian embryo provide a particularly attractive model given their capacity for long-term propagation as stem cells under defined culture conditions and their potential to generate all cell types of the adult organism (Brook and Gardner, 1997). The pluripotency of ESCs is dependent on a set of core transcriptional regulators, including Nanog, Oct4/Pou5f1 and Sox2 (NOS) (Kim, 2008; Loh et al., 2006). The co-expression of Oct4, Sox2, and Klf4, a member of a family of transcriptional regulators with redundant roles in ESC maintenance (Jiang et al., 2008), is sufficient for a broad range of differentiated cell types to acquire a pluripotent state closely resembling that of ESCs (Gonzalez et al., 2011; Takahashi and Yamanaka, 2006). Direct analysis of the targets of these transcriptional regulators has demonstrated that core pluripotency factors co-occupy cis-regulatory elements near ESC specific genes, providing strong evidence for co-regulatory inputs into the pluripotency GRN, as well as mutual reinforcement of each factor's own expression (Chen et al., 2008; Jiang et al., 2008; Kim, 2008; Loh et al., 2006; Marson et al., 2008b).

Several secreted factors are pivotal to maintaining ESC properties; their addition to culture medium replaces the requirement for serum and feeder cell support in the maintenance and propagation of ESCs (Evans and Kaufman, 1981; Williams et al., 1988; Ying et al., 2003). In particular, LIF acts through Stat3 to maintain the pluripotency of mouse ESCs (mESCs), whereas bone morphogenetic protein (BMP)-directed activation of inhibitor of DNA binding (Id) regulatory factor family members replaces serum requirements (Niwa et al., 1998; Ying et al., 2003).

Recent studies have identified two small molecule pathway modulators, PD0325901 (PD03) and CHIR99021 (CHIR), which substitute for LIF and BMP in defined ESC medium to enable the isolation and propagation of mouse ESCs (Ying et al., 2008b), and for the first time, ESCs from the rat (Li et al., 2008). PD03 is an inhibitor of mitogen-activated protein kinase kinase (MEK) (Bain et al., 2007); MEK action lies downstream of several receptor tyrosine kinase-mediated signaling pathways (McKay and Morrison, 2007) including the Fgf pathway. Fgf signaling is critical in establishing and maintaining trophectodermal (TE) precursors, the first differentiated cell lineage to be established by the totipotent mammalian embryo (Lanner and Rossant, 2010). CHIR inhibits glycogen synthase kinase-3 (GSK3); as GSK3β-directed phosphorylation and degradation of β-catenin suppresses canonical Wnt signaling, CHIR is a potent agonist of the Wnt signaling pathway (Aberle et al., 1997; Bain et al., 2007; Murray et al., 2004; Rubinfeld et al., 1996). In canonical Wnt signaling, the accumulation of cytoplasmic β-catenin enables complex formation with members of the Lef/Tcf family of DNA-binding, transcriptional regulators (Taelman et al., 2010). In the absence of β-catenin, Lef/Tcf factors bind DNA directly at a consensus Lef/Tcf site, and recruit transducin-like enhancer of split (TLE) proteins to silence target gene activity. In contrast, dimerization with β-catenin generates transcriptional activating complexes that bind to cis-regulatory modules (CRMs) activating target genes (Nusse, 2005).

Analysis of ESC culture and embryonic development provide conflicting views of the role of β-catenin-dependent, canonical Wnt signaling on ESC cultures. Together, LIF and Wnt3a, an activator of canonical Wnt signaling, are reported to support ESC pluripotency in the absence of feeder cells but in the presence of serum (Ogawa et al., 2006; Singla et al., 2006), and even in the absence of both feeder cells and serum (Ten Berge et al., 2011). Further, CHIR-mediated

stimulation of canonical Wnt signaling in the presence of PD03 blocks an intrinsic tendency of mouse ESCs to differentiate, enabling continued replication of ESCs in a pluripotent state (Ying et al., 2008b). BIO, another GSK-3 inhibitor, has been reported to maintain ESCs via up-regulation of LIF (Sato et al., 2004), and to enhance a cell fusion-mediated somatic cell reprogramming process through the accumulation of β-catenin (Lluis et al., 2008). Wnt signaling also promotes reprogramming to induced pluripotent cells (iPSCs), substituting for c-Myc in the efficient propagation of iPSCs derived from mouse embryonic fibroblasts infected with Sox2, Oct4, and Klf4 (Marson et al., 2008a). The down-regulation of "stemness marker genes" in ESCs lacking functional β-catenin supports a role for canonical Wnt signaling in maintenance of pluripotency (Anton et al., 2007), though a second study of an independently-generated β-catenin-deficient mES cell line draws a different conclusion (Lyashenko et al., 2011).

At the DNA level, genome wide interaction studies of canonical Wnt signaling effectors have largely focused around transcription factor 7 like 1 (Tcf7l1, commonly known as Tcf3), a transcriptional component that is thought to predominantly repress Wnt-target genes. Tcf3 binding shows a strong intersection at sites co-bound by major pluripotency regulators (Cole et al., 2008; Marson et al., 2008b). Recent reports indicate that a loss of Tcf3 can substitute for CHIR in 2i, which is consistent with an inhibitory action of this member on the pluripotency program (Wray et al., 2011). CHIR-mediated stimulation of β-catenin activity is proposed to both abrogate Tcf3 repression on the pluripotency network through a transactivation independent mechanism and to promote pluripotency through an interaction with Oct4 (Kelly et al., 2011; Wray et al., 2011; Yi et al., 2011). Together these data provide evidence for a canonical Wnt/β-catenin pathway action in promoting the pluripotent state of stem cells.

Conversely, canonical Wnt signaling has also been shown to induce specification of TE and mesendoderm lineages (Bakre et al., 2007; He et al., 2008; Lindsley et al., 2006), and mouse embryos lacking β-catenin, Wnt3, or two Wnt co-receptors, Lrp5 and Lrp6, arrest prior to gastrulation linking canonical Wnt signaling to axial specification and mesendodermal induction (Barrow et al., 2007; Behringer et al., 1999; Birchmeier et al., 2000; Merrill et al., 2004; Morkel et al., 2003; Skarnes et al., 2004). The conflicting reports present a mechanistic paradox: how does β-catenin-dependent canonical Wnt signaling promotes both the stem cell state and the early commitment of pluripotent cells to specific cell lineages of the gastrulating embryo?

To address this question, we engineered a mESC line to produce a Biotin-FLAG epitope tagged form of β-catenin from the β-catenin (*Ctnnb1)* locus, and performed genome-wide ChIP-seq to directly address β-catenin target sites on canonical Wnt signaling activation in mouse ESCs. When these data are viewed in conjunction with extensive expression profiling of ESCs under pluripotency and differentiation promoting conditions, together with DNA binding studies of key pluripotency determinants and their complex formation with β-catenin, a mechanistic model emerges that can reconcile the opposing actions of canonical Wnt signaling discussed above.

Results

Genome-wide profiling of the canonical Wnt regulatory network in mESCs

To take advantage of *in vivo* biotinylation and FLAG-tag technologies in analyzing canonical Wnt signaling in mESCs, we generated a *Ctnnb1-Biotin-3xFLAG* knock-in ESC line (*Ctnnb1-BioFLneo* ESC) using gene-targeting strategies (Figure 2.1A) (de Boer et al., 2003; Wang et al.,

2006). The modified allele places a carboxyl-terminal epitope tag on β-catenin comprising three tandem copies of a FLAG (3xFLAG) epitope (Hernan et al., 2000) and a short peptide that serves as a substrate for *in vivo* biotinylation in cells expressing the *Escherichia coli* biotin ligase, BirA (de Boer et al., 2003; Howard et al., 1985; Schatz, 1993) (A). Correct targeting of the modified *Ctnnb1* knocked-in allele was confirmed by long-range PCR (Figure 2.1B). *Ctnnb1-BioFLneo* ESCs were then engineered to stably express BirA to ensure biotinylation of β-catenin-BioFL proteins (*Ctnnb1-BioFLneo; BirA* ESC). We confirmed the integrity, specificity and activity of the allele through the following observations. First, production and localization of β-catenin-BioFL protein was comparable to that of the wild-type protein (Figure 2.1C and 2.1D). Second, biotinylated β-catenin-BioFL proteins were detected using streptavidin conjugated reagents in *Ctnnb1-BioFLneo;BirA* ESCs, but not in *Ctnnb1-BioFLneo* ESCs (Figure 2.1D). Finally, biotinylated β-catenin-BioFL appeared to function normally; mice homozygous for *Ctnnb1-BioFL* alleles carrying BirA ligase are viable with no apparent abnormalities (SO and APM, manuscript in preparation). As a control cell line for subsequent analyses, we also generated a BirA-expressing ESC line (*BirA* ESC; Figure 2.1E).

To better understand the roles of canonical Wnt signaling in ESC biology, we set out to identify genomic targets of β-catenin, applying ChIP-seq to *Ctnnb1-BioFLneo;BirA* ESCs. *Ctnnb1-BioFLneo;BirA* ESCs were cultured on feeder cells in standard condition with serum+LIF complete media (abbreviation: CM), and treated with CHIR (CM+CHIR) for 16 hours, then subjected to a series of ChIP-seq procedures. β-catenin-DNA complexes were pulled down using anti-FLAG antibody (FLAG-ChIP), or streptavidin (Biotin-ChIP), in parallel. DNA obtained from each ChIP procedure was independently sequenced. We also repeated FLAG-ChIP on *Ctnnb1-BioFLneo;BirA* ESCs without CHIR treatment.

Figure 2.1

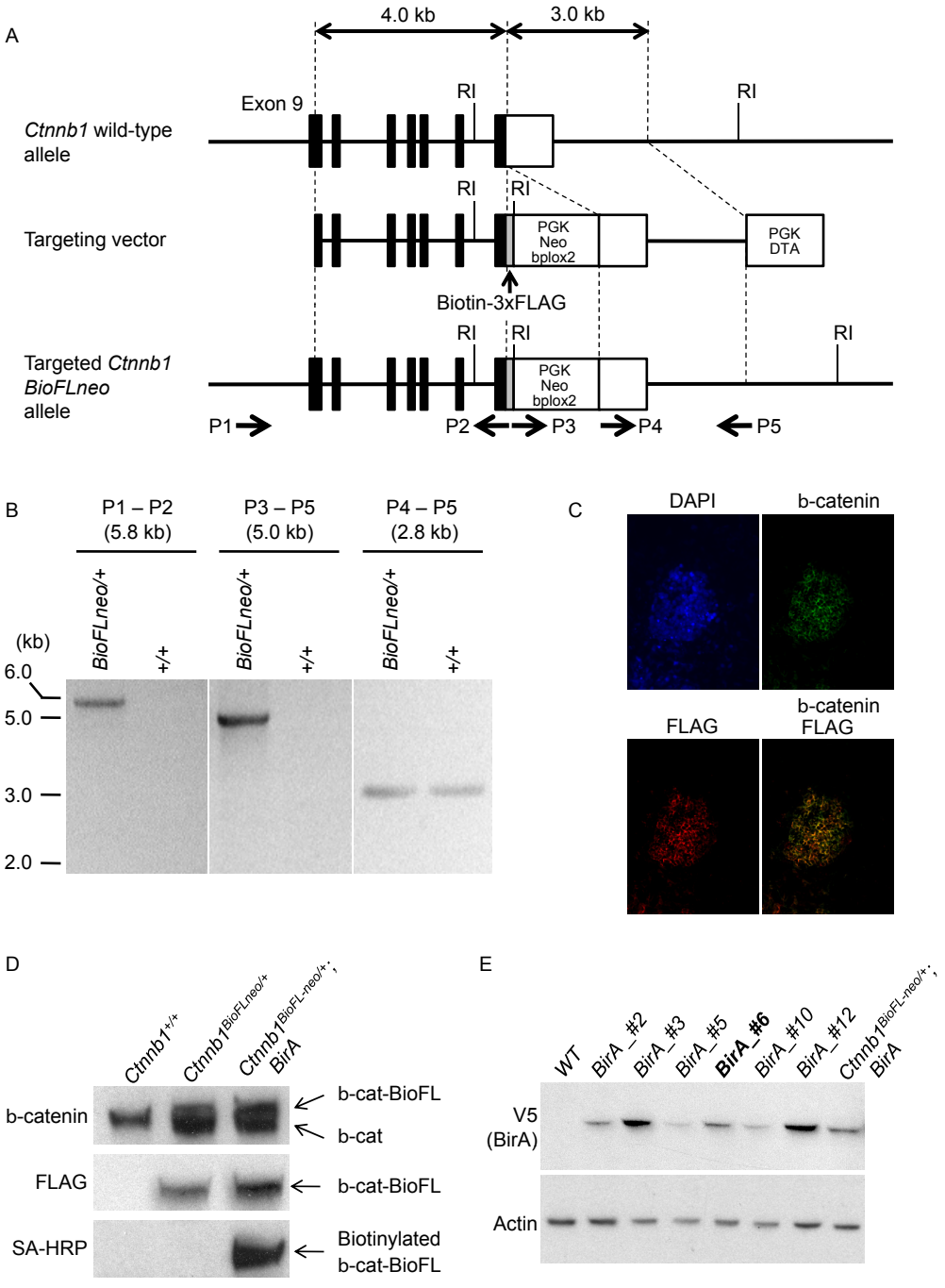

A
4.0 kb
3.0 kb
Exon 9
RI
RI
Ctnnb1 wild-type allele
RI
RI
Targeting vector
PGK Neo bplox2
PGK DTA
Biotin-3xFLAG
RI
RI
RI
Targeted Ctnnb1 BioFLneo allele
PGK Neo bplox2
P1
P2
P3
P4
P5
B
P1 – P2 (5.8 kb)
P3 – P5 (5.0 kb)
P4 – P5 (2.8 kb)
BioFLneo/+
+/+
(kb)
6.0
5.0
3.0
2.0
C
DAPI
b-catenin
FLAG
b-catenin FLAG
D
Ctnnb1+/+
Ctnnb1BioFLneo/+
Ctnnb1BioFL-neo/+; BirA
b-catenin
b-cat-BioFL
b-cat
FLAG
b-cat-BioFL
SA-HRP
Biotinylated b-cat-BioFL
E
WT
BirA_#2
BirA_#3
BirA_#5
BirA_#6
BirA_#10
BirA_#12
Ctnnb1BioFL-neo/+; BirA
V5 (BirA)
Actin

Figure 2.1 (Continued)
Generation of *Ctnnb1-BioFLneo;BirA* and *BirA* ESCs.
(A) Targeting strategy for *Ctnnb1-BioFLneo* knock-in alleles.
(B) Confirmation of targeted *Ctnnb1-BioFLneo* loci by long-range PCR. Primers used are shown in (A).
(C) Immunofluorescence for β-catenin (green) and FLAG (red) in *Ctnnb1-BioFLneo;BirA* ESC. Nucleus was stained with DAPI (blue). Bar, 50 mm.
(D) Protein expression of β-catenin (b-cat), b-cat-BioFL, and biotinylated b-cat-BioFL. SDS-PAGE and western blotting were performed using whole cell lysates of WT v6.5 ($Ctnnb1^{+/+}$), *Ctnnb1-BioFLneo* ($Ctnnb1^{BioFLneo/+}$), and *Ctnnb1-BioFLneo;BirA* ($Ctnnb1^{BioFL\text{-}neo/+}$;*BirA*) ESCs. Note that biotinylated b-cat-BioFL was detected in *Ctnnb1-BioFLneo;BirA,* neither in *Ctnnb1-BioFLneo* nor WT ESCs.
(E) Comparison of BirA expression between six clones of *BirA* ESCs, determined by SDS-PAGE and western blotting. *BirA* ESC clone #6 (bold) showed comparable expression of BirA to *Ctnnb1-BioFLneo;BirA* ESC. The clone #6 was used for a mock ChIP control.

We obtained 15947 and 16069 binding regions for Biotin-ChIP and FLAG-ChIP replicates, respectively, from ChIP-seq of CHIR treated *Ctnnb1-BioFLneo;BirA* ESCs (Figure 2.3A-C; two sample peak calling by MACS using input as controls under p-value <1e-5 and FDR<0.05; see methods for detailed peak calling program and criteria) (Zhang et al., 2008). In contrast, only a small number of regions were bound by β-catenin in CM without CHIR (data not shown), suggesting only background levels of endogenous canonical Wnt signaling in serum+LIF supplemented feeder-supported cultures.

The intersection of the two CHIR-dependent data sets identified 9885 regions (62.0% in Biotin-ChIP and 61.5% in FLAG-ChIP) in common (correlation=0.89, Figure 2.3C). Shared peaks have a higher peak ranking than ChIP regions unique to a single dataset suggesting that the intersection represents the most robust set of bone-fide interaction sites (Figure 2.3D). Representative peaks associated with pluripotency sustaining transcriptional components were validated by qPCR (Figure 2.3E). The intersection of the two datasets formed the foundation for subsequent analysis (Figure 2.2A). All binding sites were annotated relative to ref-seq gene predictions (Figure 2.2B). When compared across the genome, β-catenin associated regions show enrichment within 10kb of the transcriptional start site (TSS), and a relative depletion in intronic and exonic regions (Figure 2.2B). Approximately 16% of all annotated genes are associated with β-catenin binding in CHIR treated mESCs using a 'gene' definition as the region 10 kb upstream of the TSS plus the gene body.

Motif analysis was performed on 400 base-pair regions centered on the peak summit of β-catenin association to identify statistically enriched DNA motifs within the data set (Figure 2.3F). As expected, the DNA target site for Lef/Tcf factors, the DNA-binding partner for β-

Figure 2.2

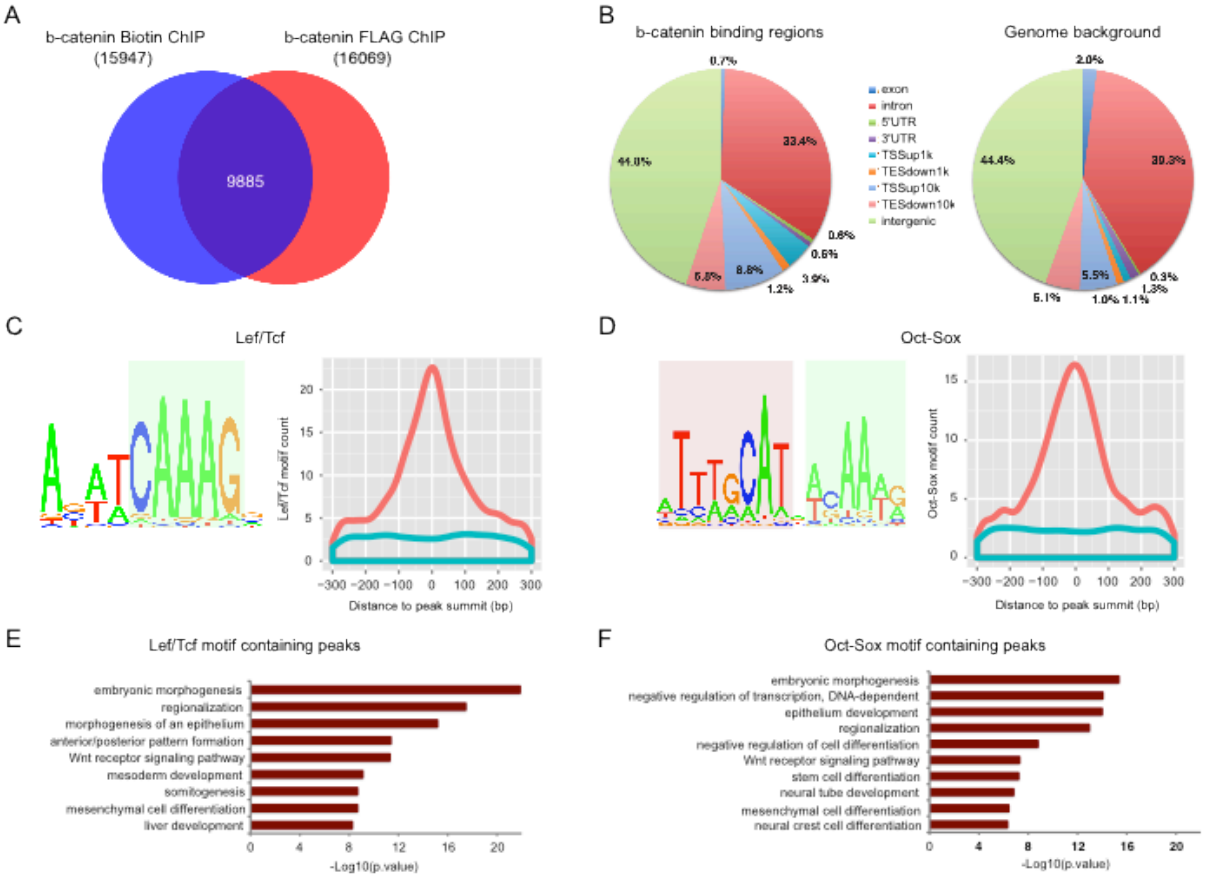

A
b-catenin Biotin ChIP
(15947)
b-catenin FLAG ChIP
(16069)
9885
B
b-catenin binding regions
Genome background
exon
intron
5'UTR
3'UTR
TSSup1k
TESdown1k
TSSup10k
TESdown10k
intergenic
33.4%
44.0%
0.8%
0.6%
8.8%
5.8%
1.2%
3.9%
2.0%
30.3%
44.4%
5.5%
0.3%
1.3%
5.1%
1.0%
1.1%
C
Lef/Tcf
Lef/Tcf motif count
Distance to peak summit (bp)
D
Oct-Sox
Oct-Sox motif count
Distance to peak summit (bp)
E
Lef/Tcf motif containing peaks
embryonic morphogenesis
regionalization
morphogenesis of an epithelium
anterior/posterior pattern formation
Wnt receptor signaling pathway
mesoderm development
somitogenesis
mesenchymal cell differentiation
liver development
-Log10(p.value)
F
Oct-Sox motif containing peaks
embryonic morphogenesis
negative regulation of transcription, DNA-dependent
epithelium development
regionalization
negative regulation of cell differentiation
Wnt receptor signaling pathway
stem cell differentiation
neural tube development
mesenchymal cell differentiation
neural crest cell differentiation
-Log10(p.value)

Figure 2.2 (Continued)
Genome-wide mapping of β-catenin binding regions in mESCs cultured in CM.

(A) Venn diagram showing the overlap between β-catenin Biotin ChIP-seq and FLAG ChIP-seq peaks.

(B) Genome-wide distribution of β-catenin binding regions relative to mouse genes compared with random control region genomic distribution. Binding regions were annotated as exon, introns, 5' un-translated region (5' UTR), 3' UTR, within 0-1 kb upstream of TSS (TSSup1k), within 1-10 kb upstream of TSS (TSSup10k), within 0-1 kb downstream of TES (TESdown1k), within 1-10kb downstream of TES (TESdown10k), or > 10kb away from the nearest genes (intergenic).

(C) (D) Top enriched motifs recovered from de novo motif analysis of β-catenin binding regions. Left panels show motif logos. HMG box motif is highlighted in light blue, and POU family motif in light red. Right panels show histogram of motifs ± 300bp around peak summit of β-catenin (orange) or matched control peak (blue).

(E) (F) GO terms enriched for β-catenin peaks containing Lef/Tcf motif (E) or Oct-Sox motif (F) using GREAT. The $-\log_{10}$ of the raw binomial p-value is reported.

Figure 2.3

A

	b-catenin Biotin	b-catenin FLAG	Input control
Total reads	32272888	37713817	52329927
Unique reads	11934863	16100920	23428242
Peaks identified (FDR<5)	15947	16069	-*

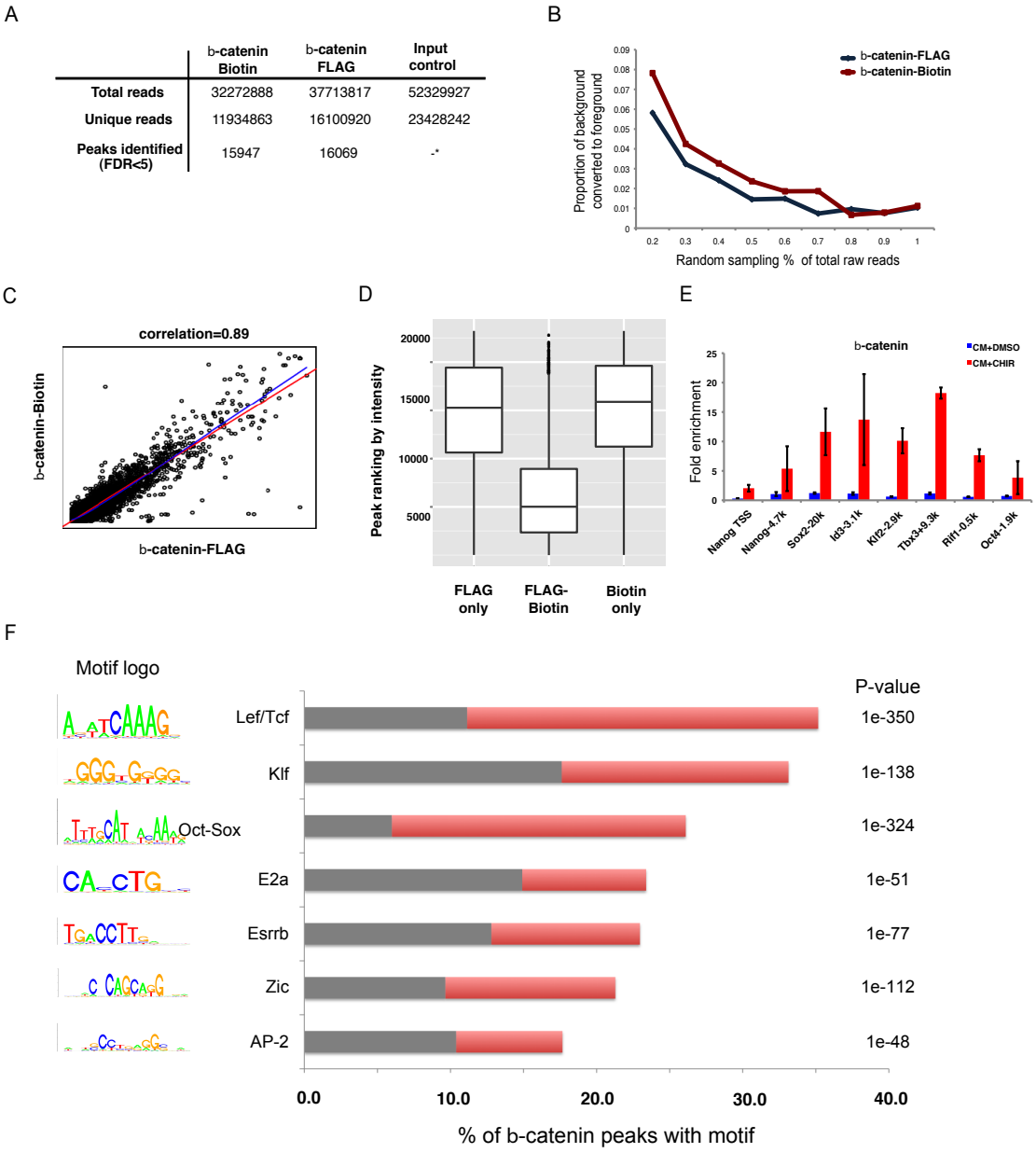


Figure 2.3 (Continued)
Characterization of β-catenin ChIP-seq data, Related to Figure 2.2.
(A) Sequencing summary of β-catenin ChIP-seq. *Reversing ChIP and control in MACS yields 784 and 342 negative peaks for β-catenin-FLAG and β-catenin-Biotin, respectively.

(B) Sequencing depth in β-catenin FLAG and Biotin ChIP-seq according to the method described in Cao et al(Cao et al., 2010).

(C) Peak signal strength correlation of the two replicates of FLAG and Biotin ChIP-seq. Each circle represents one binding peak that is called in both replicates. The X-axis and Y-Axis are $-\log_{10}$ (p-value) of each peak in the two replicates according to MACS output. The red and blue line represents regression line using linear model and LOWESS, respectively.
(D) Ranking based on peak intensity for the three categories of β-catenin peaks: FLAG only, FLAG-Biotin intersected, and Biotin only. Peaks with strong intensities rank high and therefore have small values in ranking.
(E) qPCR validation of representative β-catenin binding peaks. Data are represented as mean ± s.e.m of three independent experiments (n=3).
(F) Motifs over-represented in β-catenin-bound regions. Putative enriched motifs logos, motif occurrences in β-catenin peak regions and matched control regions (grey), and $-\text{Log}_{10}$(P-value) calculated from two-proportion z-test are shown. Motif mapping was performed using CisGenome package using likelihood ratio of 500.

catenin, was highly enriched: 35.0% of all β-catenin peaks predicted a Lef/Tcf site, versus 11.1% in matched control regions (two-proportion z test, p-value < 1e-350). Strikingly, an Oct-Sox composite motif was also highly enriched appearing in 26.1% of all β-catenin peaks (two-proportion z test, p-value < 1e 324) and like Lef/Tcf predictions, this motif was centered at the predicted peak of β-catenin binding (Figure 2.2C and 2.2D). We also identified motifs that matched binding sites for Klf4, Zic, Esrrb, E2a, and AP-2 (Figure 2.3F). The distribution of these suggests enrichment in the region but a less direct association with β-catenin binding (data not shown). Importantly, the enrichment of Lef/Tcf sites provides strong support for the quality of the data set, while the enrichment of Oct-Sox motifs near β-catenin peak summits suggests an interplay between β-catenin and the pluripotency circuit on canonical Wnt signaling stimulation. Interestingly, gene ontology (GO) analysis using Genomic Regions Enrichment of Annotations Tool (GREAT)(McLean et al., 2010) revealed that while embryogenesis-related and Wnt receptor signaling pathway-related genes were both enriched in Lef/Tcf motif-containing and Oct-Sox motif-containing β-catenin peaks, the former category was also enriched in mesoderm development-related term, and the latter stem cell- and neural-related terms (Figure 2.2E and 2.2F).

Analysis of β-catenin, Tcf3, Sox2, Oct4, and Nanog interactions at target genes points to distinct enhancer modules mediating the actions of canonical Wnt signaling

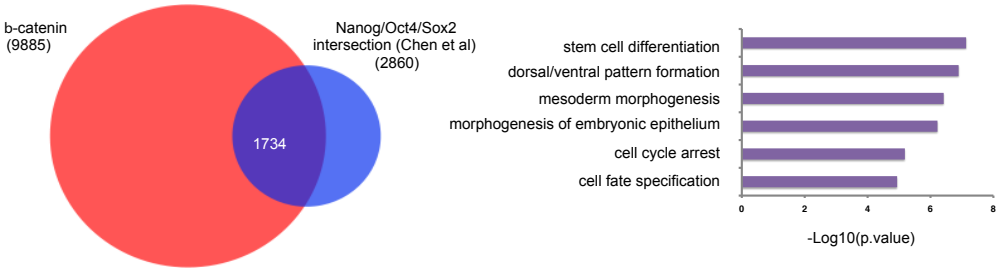
The recovery of Oct-Sox motif within β-catenin binding regions prompted us to compare the β-catenin data with previously-published ChIP-seq data for 19 TFs associated with maintenance of pluripotency, induction of iPSCs, and Wnt action (Figure 2.4A): the core pluripotency factors

Figure 2.4

A

Factor	Total reads	Unique reads	Peaks identified (p.value<1e-5)	Factor	Total reads	Unique reads	Peaks identified (p.value<1e-5)
Tcf3 (Marson)	15809150	8021327	10973	Zfx (Chen)	17968941	6709856	34599
Nanog (Marson)	17802070	8206266	27088	E2f1 (Chen)	30489540	10503301	31225
Pou5f1 (Marson)	8763614	4507625	32768	C-myc (Chen)	26967738	11050661	8835
Sox2 (Marson)	25437792	10537485	29039	N-myc (Chen)	23180904	7465082	15604
Nanog (Chen)	24203113	11734402	19772	Ezh2 (Ku)	17008829	7342512	5408
Pou5f1 (Chen)	24021520	7828613	9956	Ring1b (Ku)	7877125	3735193	3024
Sox2 (Chen)	23073668	7624108	9977	Suz12 (Chen)	25553001	9841465	4971
Smad1 (Chen)	20166436	6500938	1713	p300 (Creyghton)	15215665	8148619	30734
Esrrb (Chen)	25879362	11471789	87576	H3K4me1 (Creyghton)	7616941	6495114	13163
Nr5a2 (Heng)	16352448	8178284	8149	H3K27ac(1) (Creyghton)	14657880	10531543	25739
Stat3 (Chen)	21995262	8166187	7955	H3K27ac(2) (Creyghton)	18002174	11076935	31930
Tcfcp2l1 (Chen)	31573848	11154033	48147	H3K4me2 (Meissner)	22955254	6935116	50602
Tbx3 (Han)	19051474	6933334	268	H3K4me3 (Mikkelsen)	19618320	9312859	25320
Klf4 (Chen)	22126359	6699589	28734	H3K27me3 (Mikkelsen)	17748450	8067660	10821
CTCF (Chen)	24319003	7424969	55694				

B



C

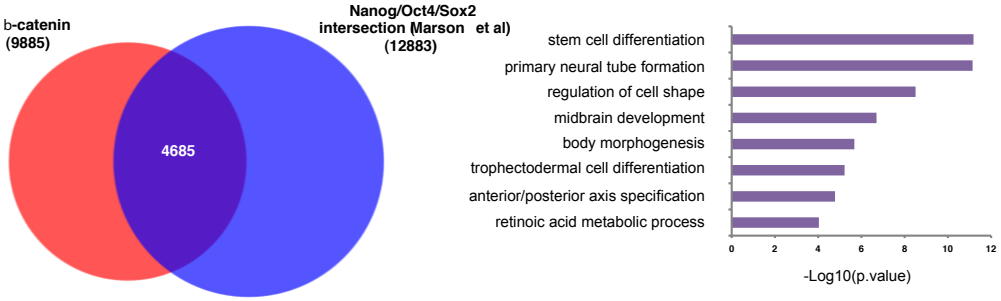


Figure 2.4 (Continued)
Summary for mESC ChIP-seq data from literature.
(A) Summary of published ChIP-seq data used in this study.
(B and C) Venn diagram showing intersection of β-catenin binding regions with Nanog/Oct4/Sox2 ones obtained from two separate studies (Chen et al., 2008; Marson et al., 2008b). GO terms enriched for genes near the intersected peaks are listed.

Nanog/Oct4/Sox2 (NOS, see Figure 2.4B and 2.4C for the comparison of two independent datasets) (Chen et al., 2008; Marson et al., 2008b); Smad1/Stat3, effectors of key mESCs signaling pathways (Chen et al., 2008); Tcfcp2l1/Tbx3/Klf4/C-Myc/N-Myc/Zfx, reprogramming factors important for self-renewal (Chen et al., 2008; Han et al., 2010); Ring1b/Ezh2/Suz12, components of polycomb repressive complexes (PRC) (Chen et al., 2008; Ku et al., 2008); Esrrb/Nr5a2, nuclear receptors linked to the ESC state (Chen et al., 2008; Heng et al., 2010); and Tcf3, the most abundant of the Lef/Tcf family of canonical Wnt transcriptional effectors in mESCs (Marson et al., 2008b).

Through pair-wise co-binding analyses, we were able to classify binding patterns for these regulatory factors into several clusters; notably Tcf3, Nanog, Sox2, Smad1, and Oct4 interactions most closely resembled those observed through β-catenin ChIP-seq (Figure 2.5A). Given that β-catenin regulates gene expression through Tcf TFs, of which Tcf3 is most abundantly expressed in the mESCs, we performed a two-way intersection of β-catenin and Tcf3 binding peaks taking only the β-catenin::Tcf3 regions to increase the credibility of binding events. A further intersection with NOS peaks, produced Group-A (β-catenin::Tcf3) and Group-B (β-catenin::Tcf3::NOS) (Figure 2.5B): comparison of these two categories provides an insight into whether canonical Wnt signaling action differs in the presence of NOS.

Motif enrichment, chromatin state, and functional properties of predicted target genes adjacent to these binding regions were explored in each peak dataset grouping. A clear consensus Lef/Tcf motif was the most over-represented motif in Group-A (p-value < 1e-336, two-proportion z test) (Figure 2.5C), while the most enriched motif in Group-B closely resembled the published Oct-Sox motif (p-value < 1e-561, two-proportion z test) (Figure 2.5D). Stem cell- and ectoderm-related terms were enriched in Group-B targets, while axis

Figure 2.5

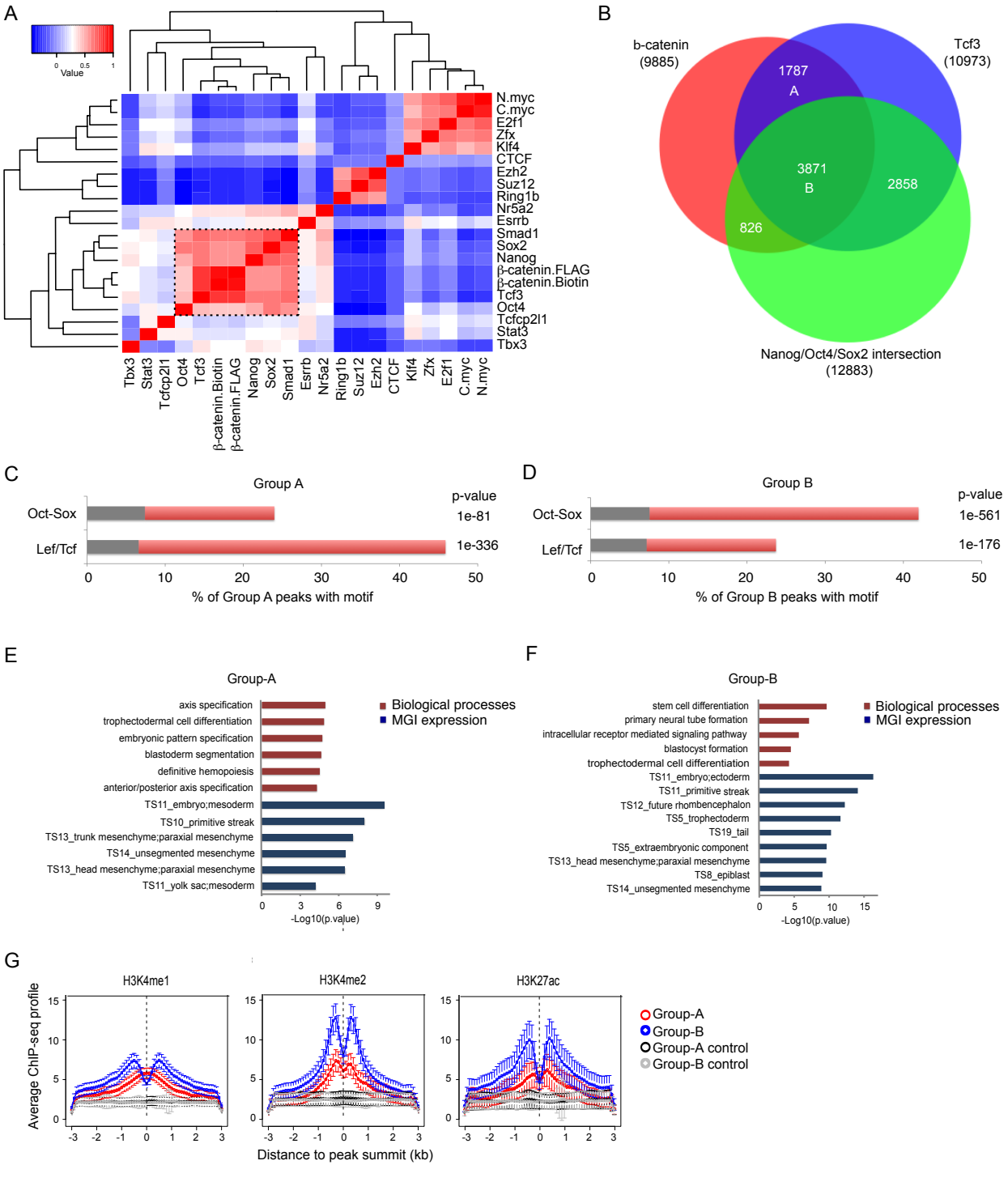

A
Value
N.myc
C.myc
E2f1
Zfx
Klf4
CTCF
Ezh2
Suz12
Ring1b
Nr5a2
Esrrb
Smad1
Sox2
Nanog
β-catenin.FLAG
β-catenin.Biotin
Tcf3
Oct4
Tcfcp2l1
Stat3
Tbx3
B
b-catenin
(9885)
Tcf3
(10973)
1787
A
3871
B
2858
826
Nanog/Oct4/Sox2 intersection
(12883)
C
Group A
p-value
Oct-Sox
1e-81
Lef/Tcf
1e-336
% of Group A peaks with motif
D
Group B
p-value
Oct-Sox
1e-561
Lef/Tcf
1e-176
% of Group B peaks with motif
E
Group-A
Biological processes
MGI expression
axis specification
trophectodermal cell differentiation
embryonic pattern specification
blastoderm segmentation
definitive hemopoiesis
anterior/posterior axis specification
TS11_embryo;mesoderm
TS10_primitive streak
TS13_trunk mesenchyme;paraxial mesenchyme
TS14_unsegmented mesenchyme
TS13_head mesenchyme;paraxial mesenchyme
TS11_yolk sac;mesoderm
-Log10(p.value)
F
Group-B
Biological processes
MGI expression
stem cell differentiation
primary neural tube formation
intracellular receptor mediated signaling pathway
blastocyst formation
trophectodermal cell differentiation
TS11_embryo;ectoderm
TS11_primitive streak
TS12_future rhombencephalon
TS5_trophectoderm
TS19_tail
TS5_extraembryonic component
TS13_head mesenchyme;paraxial mesenchyme
TS8_epiblast
TS14_unsegmented mesenchyme
-Log10(p.value)
G
H3K4me1
H3K4me2
H3K27ac
Average ChIP-seq profile
Distance to peak summit (kb)
Group-A
Group-B
Group-A control
Group-B control

Figure 2.5 (Continued)
Characterization of β-catenin and ESC pluripotency factors binding.
(A) Heat map depicting the correlation of β-catenin and ESC factors bindings. Red: positive correlation; blue: negative correlation.
(B) Venn diagram of β-catenin, Tcf3, and intersection of Nanog/Oct4/Sox2 peak regions. Two groups of peaks are highlighted: Group-A: β-catenin::Tcf3, and Group-B: β-catenin::Tcf3::NOS.
(C) (D) Enriched motifs in Group-A and Group-B. Red: motif occurrence in β-catenin peaks; grey: motif occurrence in matched control regions with the same coverage. P-value was calculated according to two proportion z-test.
(E) (F) Functional annotation of Group-A and Group-B regions using GREAT. The $-\log_{10}$ of the raw binomial p-value is shown.
(G) Aggregation plots of H3K4me1, H3K4me2, and H3K27ac signals ±3 kb around the peak summit for binding regions in Group-A (red) and Group-B (blue) as well as corresponding matched control regions with standard error bars (black and grey). The analysis is done using HOMER (Heinz et al., 2010). Bin size 100 bp.

specification and mesoderm terms were over-represented in Group-A targets (Figure 2.5E and 2.5F). In terms of chromatin state, both groups displayed a strong H3K4me2 signature, an indicator of poised or active enhancer regions (He et al., 2010), but the signature was more prominent amongst Group B regions, suggestive of a more active state in ESCs (Figure 2.5G). Consistent with this view, Group-B displayed a stronger H3K4me1 and H3K27ac active enhancer signature than Group-A (Figure 2.5G) (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner et al., 2011).

In summary, the data suggest two types of β-catenin associated regulation of two distinct categories of target genes, one through a Lef/Tcf-mediated DNA interaction of β-catenin with poised enhancers around differentiation-related genes (Group A) and one through cooperative interactions of NOS and β-catenin/Lef/Tcf with active enhancers around stem cell- and ectoderm-related genes (Group B). A CisGenome browser screenshot of representative genes of Group-A (*Cdx2*, Figure 2.6A; *Axin2*, Figure 2.6C), and Group-B (*Nanog*, Figure 2.6B) shows the relative signal intensity of β-catenin to Nanog/Oct4/Sox2 data for Group-A versus Group-B associated genes.

Activation of canonical Wnt signaling directs early mesoderm differentiation

To connect DNA association profiles of these factors with canonical Wnt signaling-mediated gene expression, we intersected β-catenin ChIP peaks with neighboring genes, with a focus on those genes that displayed differential expression between mESCs cultured in CM+CHIR and CM+XAV939 (XAV); XAV is a tankyrase inhibitor antagonizing Wnt signaling (Huang et al., 2009b). Among the intersected gene set, genes associated with a canonical Wnt signaling response, and early mesoderm and TE cell types, were the most significant up-regulated gene

Figure 2.6

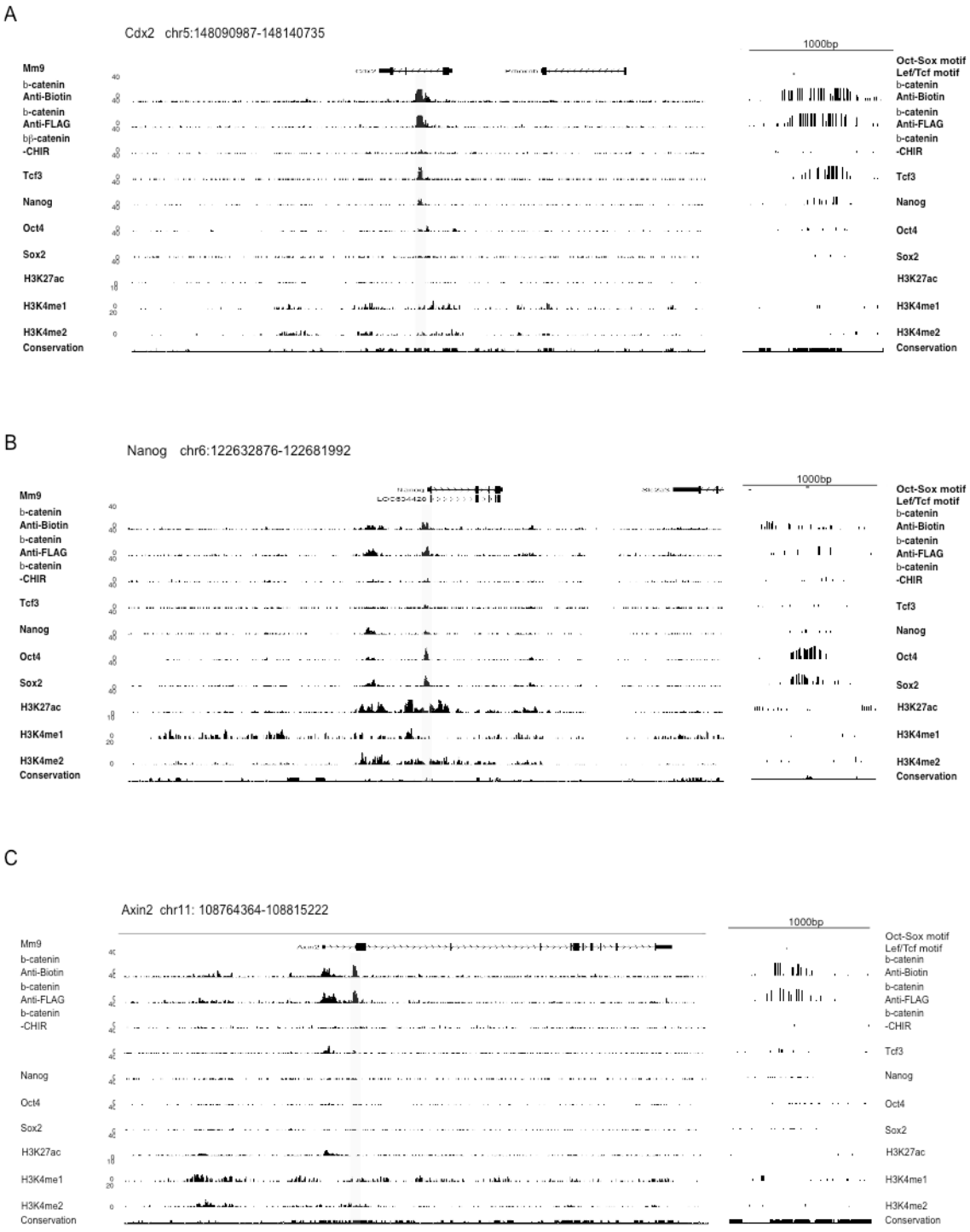

A
Cdx2 chr5:148090987-148140735
1000bp
Mm9
b-catenin
Anti-Biotin
b-catenin
Anti-FLAG
bβ-catenin
-CHIR
Tcf3
Nanog
Oct4
Sox2
H3K27ac
H3K4me1
H3K4me2
Conservation
Oct-Sox motif
Lef/Tcf motif
B
Nanog chr6:122632876-122681992
1000bp
C
Axin2 chr11: 108764364-108815222
1000bp

Figure 2.6 (Continued)
CisGenome browser screenshots showing combinatorial binding pattern of β-catenin and core pluripotency factors in CM.
β-catenin binding to known Wnt target genes related to differentiation (*Cdx2*), pluripotency (*Nanog*, B), and canonical Wnt target gene (*Axin2,* C). Endogenous association of β-catenin is also displayed in the absence of CHIR stimulation (CHIR-). Tcf3, Nanog, Oct4, Sox2 and histone modification profiles displayed here are from published datasets (see Results). A zoom in on highlighted regions with motif annotation is displayed to the right side.

targets. Even though *Oct4* and *Sox2* were diminished upon CHIR treatment (see above), pluripotency-related genes showed no strong differential expression.

The integrity of our microarray data set was supported by two analyses: prediction of canonical Wnt signaling target genes and correlation with published expression data sets. First, we applied an approach based upon an empirical finding that the potential of a gene being a direct target for a given TF decreases monotonically as a function of the distance of the binding site to that gene's TSS (Tang et al., 2011). Second, using a rank product based method, we made a probabilistic prediction of β-catenin target gene list (Figure 2.7A) (Breitling et al., 2004). Using a false discovery rate (FDR) of 10%, we obtained 376 and 362 putative direct target genes that were positively and negatively regulated by CHIR, respectively, in CM. This method accurately predicted known target genes of canonical Wnt signaling, such as *Axin2*, *T*, *Sp5*, *Lef1*, *Cdx2*, and *Tcfcp2l1*. Interestingly, the identification of *Porcn*, which encodes a key factor in the palmitoylation and secretion of Wnt ligands (Takada et al., 2006), suggests a hitherto unrecognized positive feedback loop in canonical Wnt signaling.

We calculated the correlation of the up-regulated genes in our data with the published microarray data comparing gene expression between ESCs, mesendoderm cells (MECs), and neural ectoderm cells (NECs) (Shen et al., 2008). The top 100 genes displaying a high MEC/ESC expression ratio showed some correlation (> 0.5) with genes exhibiting a high CHIR/XAV expression ratio. No correlation was observed with genes associated with neural ectoderm development (high NEC/ESC) (Figure 2.7B), consistent with the known role of canonical Wnt signaling in inducing mesendoderm and suppressing neural ectoderm development (Barrow et al., 2007; Takada et al., 1994; Yamaguchi et al., 1999).

Figure 2.7

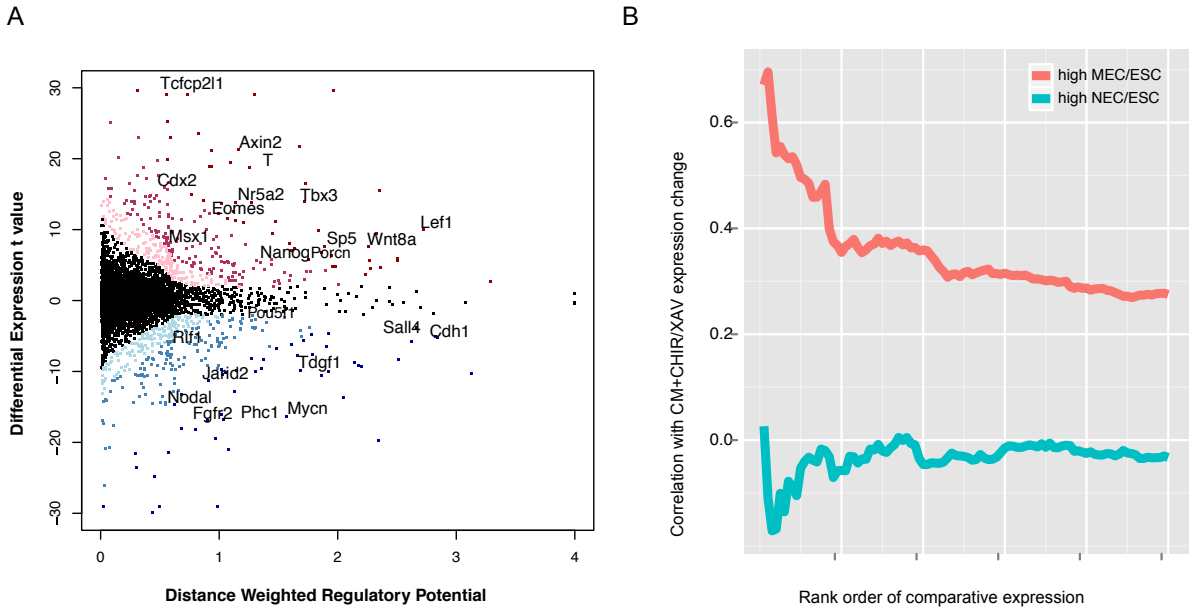

A
B
Tcfcp2l1
Axin2
T
Cdx2
Nr5a2
Tbx3
Eomes
Lef1
Msx1
Sp5
Wnt8a
Nanog
Porcn
Pou5f1
Sall4
Cdh1
Rif1
Tdgf1
Jarid2
Nodal
Fgfr2
Phc1
Mycn
Differential Expression t value
Distance Weighted Regulatory Potential
high MEC/ESC
high NEC/ESC
Correlation with CM+CHIR/XAV expression change
Rank order of comparative expression

Figure 2.7 (Continued)
Integration of β-catenin ChIP-seq and expression profiling in mESCs treated with an activator or inhibitor of canonical Wnt signaling.
(A) Scatter plot of β-catenin direct target gene prediction based on distance weighted regulatory potential score from ChIP-seq and t-value of differential CHIR/XAV expression in CM. Red dots: up-regulated genes with FDR < 0.10; blue dots: down-regulated genes with FDR < 0.10. The darker red/blue represents the higher likelihood for a gene being β-catenin direct target. The horizontal and vertical histograms reflect the distribution of the index for distance weighted regulatory potential and differential expression t value, respectively.
(B) Correlation of top 1000 genes of high MEC/ESC or NEC/ESC expression ratio (from microarray data in Shen et al., 2008) (x-axis) with their differential expression fold changes in CM+CHIR/CM+XAV (y-axis). Step size: 20 genes.

Similarity of β-catenin chromatin binding between CM+CHIR and 2i

To examine the interactions of β-catenin and pluripotency network components under the 2i conditions, *Ctnnb1-BioFLneo;BirA* ESCs were cultured in 2i medium supplemented with LIF; LIF addition enhances clonal propagation of 2i cultures (Ying et al., 2008b). ChIP-qPCR for Sox2, Oct4, and Tcf3 binding were conducted at *cis*-elements near pluripotency genes in 2i, 2i+LIF and CM conditions; binding of the three factors were comparable under all three conditions (Figure 2.8A, p-value > 0.05 from two sample t-test).

To understand which gene category was regulated by each 2i component, and how β-catenin may participate in this regulation, we performed ChIP-qPCR for β-catenin in 2i-adapted *Ctnnb1-BioFLneo;BirA* ESCs cultured with DMSO, PD03, CHIR, 2i, or 2i+LIF for 24 hours (Figure 2.8B). A strong enrichment of β-catenin was observed in CHIR-, 2i-, and 2i+LIF-treated cells at the same pluripotency-related gene regions as those bound by β-catenin in CM+CHIR (Figure 2.8C, upper panel); binding was dependent on the activation of canonical Wnt signaling as binding was lost within 24-hour of CHIR removal (Figure 2.8C; DMSO and PD03 in upper panel). Thus, stabilization of β-catenin leads to similar interactions at the DNA level in quite different culture regimens, and β-catenin likely contributes to expression of pluripotency-related genes under 2i conditions by the direct association with CRMs governing expression of the pluripotency network. Interestingly, inhibition of Mek activity by PD03 did not affect β-catenin interaction at putative regulatory regions around target genes associated with differentiation of ES cells in response to canonical Wnt signaling (Figure 2.8C, lower panel).

Figure 2.8

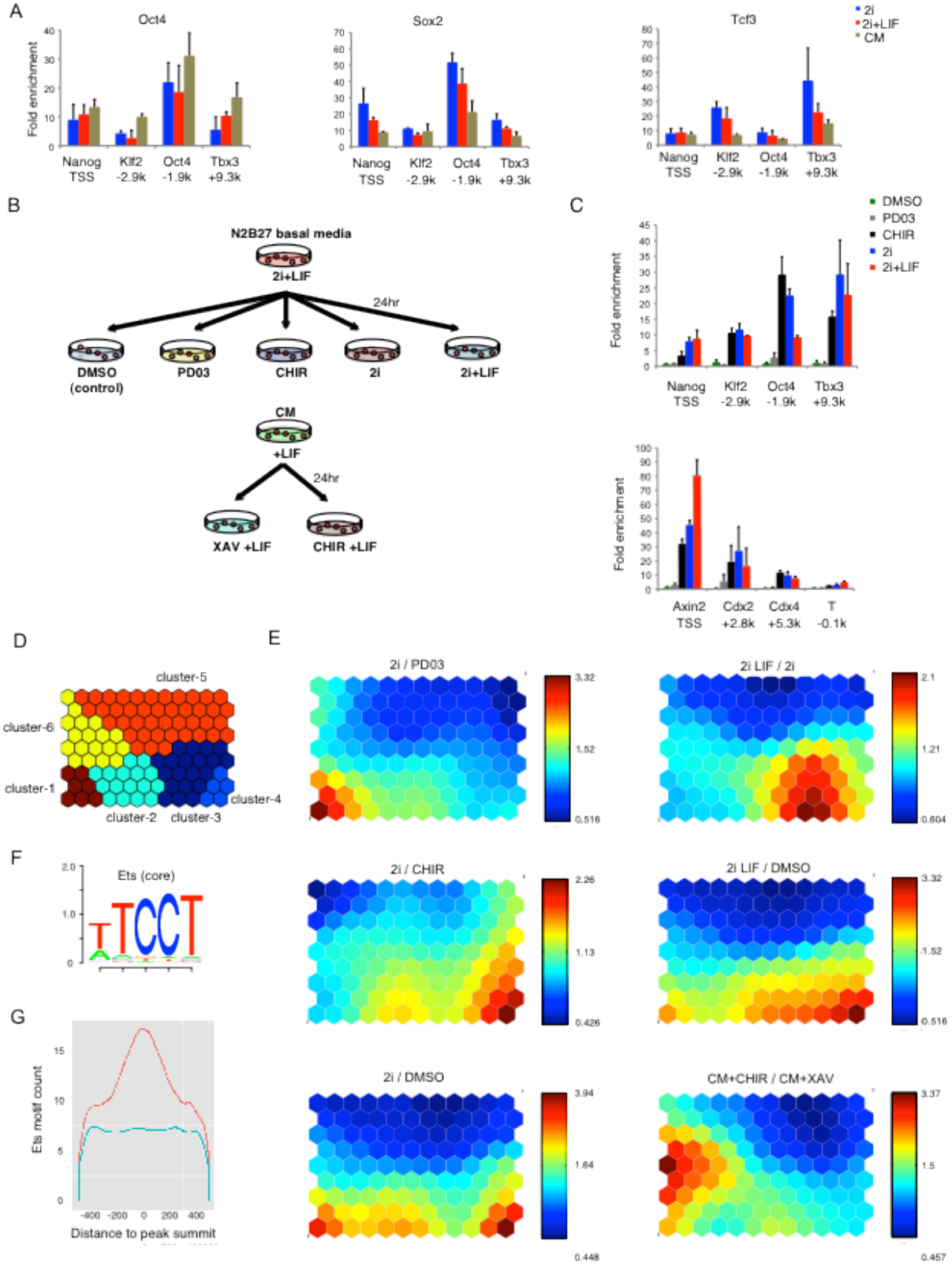

A
Oct4
Sox2
Tcf3
2i
2i+LIF
CM
Fold enrichment
Nanog TSS
Klf2 -2.9k
Oct4 -1.9k
Tbx3 +9.3k
B
N2B27 basal media
2i+LIF
24hr
DMSO (control)
PD03
CHIR
2i
2i+LIF
CM
+LIF
24hr
XAV +LIF
CHIR +LIF
C
DMSO
PD03
CHIR
2i
2i+LIF
Axin2 TSS
Cdx2 +2.8k
Cdx4 +5.3k
T -0.1k
D
cluster-5
cluster-6
cluster-1
cluster-4
cluster-2
cluster-3
E
2i / PD03
2i LIF / 2i
2i / CHIR
2i LIF / DMSO
2i / DMSO
CM+CHIR / CM+XAV
F
Ets (core)
G
Ets motif count
Distance to peak summit

Figure 2.8 (Continued)
Roles of small molecules CHIR and PD03 in 2i.
(A) ChIP-qPCR for Oct4, Sox2 and Tcf3 interaction at defined regulatory regions surrounding pluripotency target genes in mESCs cultured in 2i, 2i+LIF, and CM. Data represent the mean of biological replicates.
(B) Experimental scheme for studying the role of CHIR and PD03 in CM and 2i-adapted mESCs. ChIP-qPCR and microarray analysis were performed in 2i-adapted mESCs cultured for 24 hours with DMSO (control), PD03, CHIR, 2i, and 2i+LIF. Cells at passage 20 under the 2i+LIF condition were subjected to each assay. Cells that had been maintained in CM on feeder cells were cultured for 24 hours in CM with CHIR or XAV prior to microarray analysis.
(C) ChIP-qPCR for β-catenin at selected loci near pluripotency-related genes (upper), differentiation-related and Wnt target genes (lower) on 2i-adapted mESCs. ChIP using anti-FLAG antibodies was performed according to the experimental scheme described in (B). Data show the mean and standard error of the mean (s.e.m.) for three biological replicates.
(D) K-means clustering was used to classify genes with expression fold change > 2 in at least one comparison group of 2i/PD03, 2i/CHIR, 2i/DMSO, 2iLIF/2i, 2iLIF/DMSO, and CM+CHIR/CM+XAV. A total of 388 genes were clustered into six clusters.
(E) Individual component maps are shown for each pair-wise comparison. Top left: 2i/PD03; middle left: 2i/CHIR; bottom left: 2i/DMSO; top right: 2iLIF/2i; middle right: 2iLIF/DMSO; bottom right: CM+CHIR/CM+XAV. In general, red indicates up-regulation and blue down-regulation. The number by each color bar indicates the fold change.
(F) Five bp core Ets motif logo TCCTW from TRANSFAC motif M00339.
(G) Enrichment of Ets core motif ± 500bp around β-catenin peak summit (orange) compared with matched control regions (blue).

Differentiation-associated genes fail to be up-regulated by CHIR in 2i medium

The predominant activation of a mesendoderm lineage differentiation program by CHIR in CM (Figure 2.7B), and the association of β-catenin with differentiation-related gene regions in 2i (Figure 2.8C) raises a question of how differentiation is inhibited in 2i conditions We performed microarray analysis in 2i-adapted *Ctnnb1-BioFLneo;BirA* ESCs cultured with DMSO, PD03, CHIR, 2i, and 2i+LIF according to the same experimental scheme as in Figure 2.8B. In this, the effect of CHIR is distinguished by comparing the relative gene expression level in the ESCs cultured in 2i to those cultured in PD03 where CHIR is absent (2i / PD03). Comparing CM+CHIR over CM+XAV (CM+CHIR / CM+XAV) provides another metric of CHIR activity, while a comparison of 2i+LIF over 2i (2iLIF / 2i) sheds light on any direct effect of LIF on gene expression.

Using Kohonen Self-Organizing Maps (SOM) (Kohonen, 2012), we identified groups of genes that shared similar patterns of expression changes amongst the groups (Figure 2.8D), and visualized target gene expression change in a series of heat maps (Figure 2.8E). Genes that showed a greater than 2-fold differential expression in at least one comparison were clustered (Figure 2.8D): the six cluster associated gene names are presented in Figure 2.9. In Figure 2.8D, each hexagonal map unit represents a SOM node, which has an underlying vector of a pair-wise fold change under different conditions. In Figure 2.8E, the gradient of each pair-wise fold change is displayed separately while maintaining the same topological structure as in Figure 2.8D. On comparing Figure 2.8D and 5E, clusters-1,-3, and -4 represented genes strongly up-regulated by CHIR (2i / PD03), LIF (2i LIF / 2i), and PD03 (2i / CHIR), respectively, in 2i. Cluster-2 represented genes moderately up-regulated by CHIR (2i / PD03), LIF (2i LIF / 2i) and PD03 (2i /

Figure 2.9

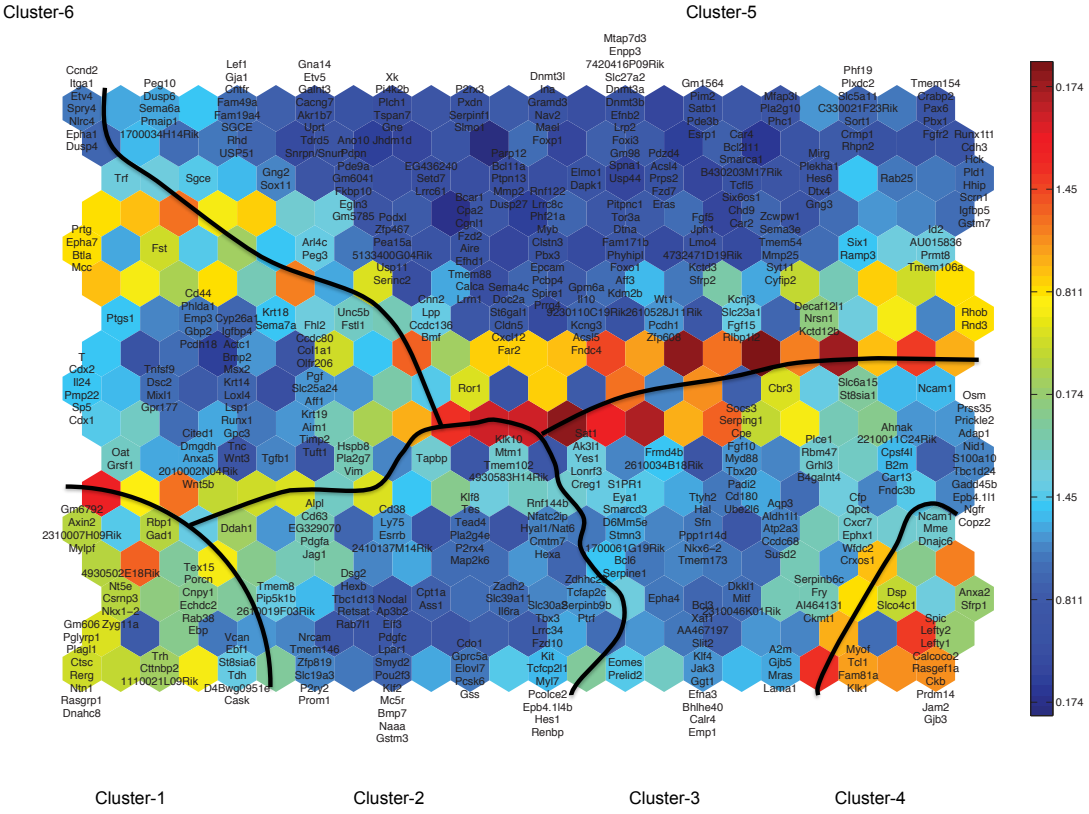

Cluster-6
Cluster-5
Cluster-1
Cluster-2
Cluster-3
Cluster-4
0.174
1.45
0.811
0.174
1.45
0.811
0.174

Figure 2.9 (Continued)
Kohonen U-matrix showing geometrical relationships between genes, representing similarity of gene expression patterns across different conditions.
In comparison with Figure 2.8D and Figure 2.8E, addition hexagons are inserted in between the original hexagons, which represent the distance between the two neighboring hexagons. A dark red color represents that the neighboring hexagons (ie, genes) are far apart in expression patterns, while a dark blue indicates that the difference of expression pattern for the neighboring hexagons is small. The color indicates the average distance between one location and the neighboring hexagons. K-means clustering into six clusters from Figure 2.8D is annotated by dividing the U-matrix with black lines.

CHIR). Genes in cluster-6 were up-regulated by CHIR in serum conditions (CM+CHIR / CM+XAV), and those in cluster-5 were down-regulated in all comparison groups.

There are three key insights from this analysis. First, CHIR, PD03, and LIF up-regulated different sets of genes in 2i, as shown in cluster-1, cluster-3, and cluster-4, respectively. Second, CHIR up-regulated different sets of genes in 2i from those in serum conditions (cluster-1 versus cluster-6): several differentiation-related genes, including T, Cdx2, Cdx1 appear in cluster-6 but were down-regulated in 2i compared to CHIR (see 2i / CHIR) (Figure 2.9). In support of the former, an Ets motif was significantly enriched in β-catenin binding regions – Ets factors provide the transcriptional output to MEK/ERK signaling (Figure 2.8F and 5G; see Discussion). Thirdly, the effect of 2i over DMSO (2i / DMSO, left bottom) appears to be an additive result of the effect of CHIR (2i / PD03, left top) and PD03 (2i / CHIR, left middle). Similarly, the effect of 2i+LIF over DMSO (2i LIF / DMSO, right middle) appears to be an additive result of CHIR (2i / PD03, left top), PD03 (2i / CHIR, left middle), and LIF (2i LIF / 2i, right top).

β-catenin complexes with Oct4 and Tcf3 at Oct-Sox motifs in 2i cultured mESCs

Binding of Oct4 and Sox2 to pluripotency-related gene regions was unaltered under 2i condition in contrast to the effect of CHIR on cells grown in CM (see earlier). To investigate the physical interaction of β-catenin and Tcf3 with these two pluripotency determinants we performed co-immunoprecipitation (Co-IP) analysis with nuclear extracts from mESCs cultured in 2i or 2i+LIF with antibodies specific to β-catenin, Tcf3, Sox2 and Oct4. Oct4 and β-catenin were co-immunoprecipitated with anti-Oct4 antibodies (Figure 2.10A); the failure of a reciprocal IP likely reflects epitope masking of the epitope recognized by the β-catenin antibody in an Oct4 complex (Kelly et al., 2011). Tcf3 was associated with β-catenin and Oct4, but not Sox2 (Figure

2.10A). Consistently, Sox2 only pulled down Oct4, not β-catenin or Tcf3. These results suggest that two regulatory complexes exist under 2i conditions: one containing β-catenin, Tcf3, and Oct4, and another with Sox2 and Oct4.

Given the overlap in ChIP peaks between Oct4, Sox2, and Tcf3, the Oct-Sox composite motif recovered in β-catenin ChIP-seq data, sequence similarity within this consensus motif between Sox and Tcf DNA binding sites, and their structural similarities sharing an HMG DNA binding domain, we tested whether Tcf3 directly bound to the Oct-Sox composite motif *in vitro* by electrophoretic mobility shift assay (EMSA). Nuclear extracts (NE) were prepared from 293T cells over-expressing *Pou5f1* (Oct4), *Sox2*, or *Tcf7l1* (Tcf3) (Figure 2.10C) and tested for their ability to bind a double-stranded DNA probe incorporating an Oct-Sox composite motif located within a distal *Pou5f1* enhancer (Figure 2.10B).

Oct4 and Sox2 alone complex with the Oct/Sox probe (Figure 2.10D, lanes 2-7) whereas no binding was observed for Tcf3 (Figure 2.10D, lanes 8-10). In contrast, Tcf3 was complexed with a Lef/Tcf binding motif (LT probe) under the same conditions (Figure 2.10D, lanes 12-20; band E). Next, we examined co-binding for cooperative interactions. Oct4 and Sox2 co-binding led to additional band (C) not seen when only Oct4 (A) or Sox2 (B) bound (Figure 2.10E, lane 2-8 (Chew et al., 2005)). As above, no Tcf3 interaction was observed with the Oct/Sox motif and Tcf3 failed to compete with Sox2 at the Sox2 binding site (Figure 2.10E, lanes 9-13). However, in the presence of Oct4 and Tcf3, additional bands were observed that likely reflect ternary complexes of Oct4 and Tcf3, with Tcf3 bound at the Sox2 site (band D in Figure 2.10E, lanes 14-22).

Figure 2.10

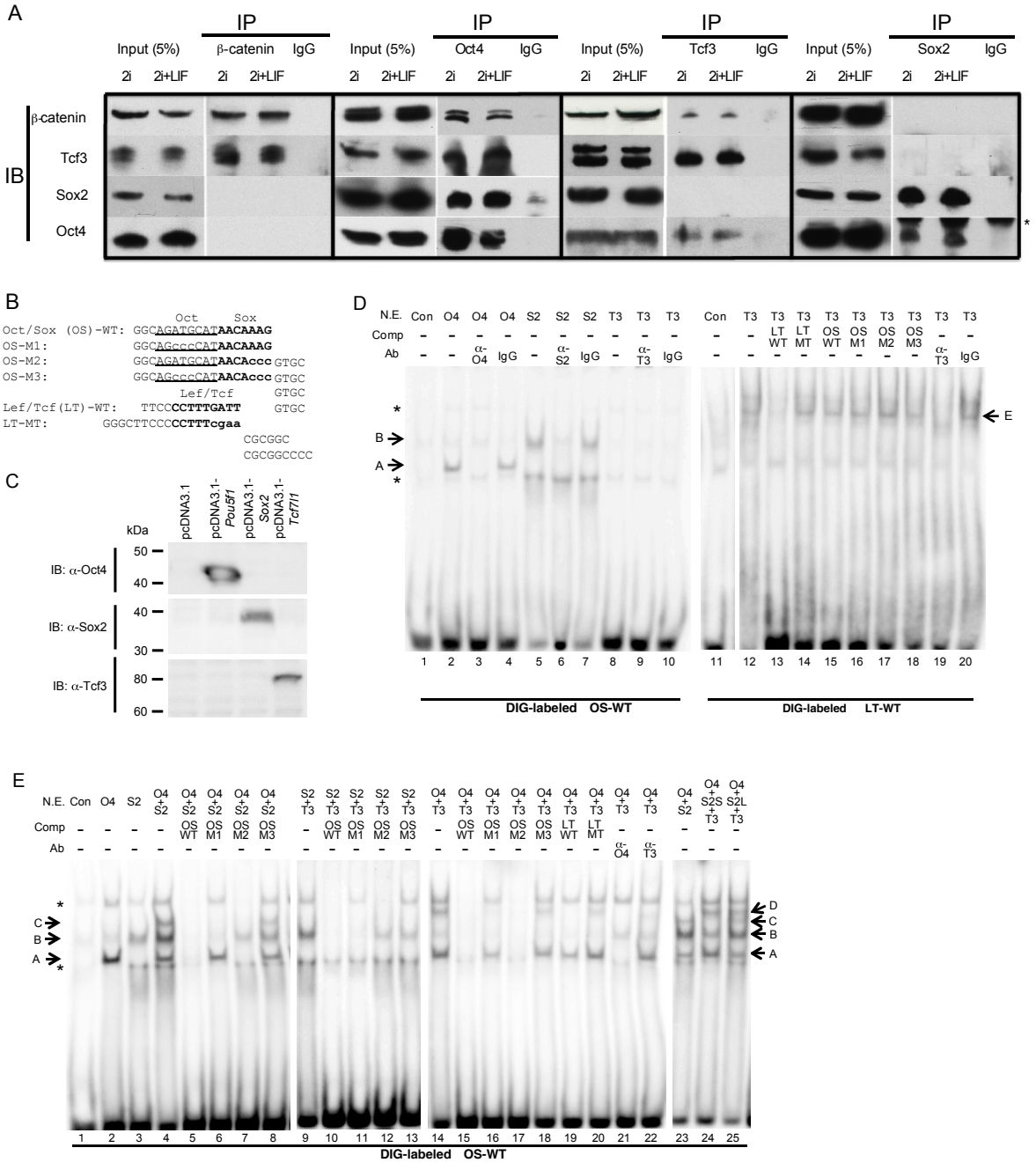

A
IP
Input (5%)
β-catenin
IgG
Oct4
Tcf3
Sox2
2i
2i+LIF
IB
β-catenin
Tcf3
Sox2
Oct4
*
B
Oct
Sox
Oct/Sox (OS)-WT: GGCAGATGCATAACAAAG
OS-M1: GGCAGcccCATAACAAAG
OS-M2: GGCAGATGCATAACAccc GTGC
OS-M3: GGCAGcccCATAACAccc GTGC
Lef/Tcf GTGC
Lef/Tcf(LT)-WT: TTCCCCTTTGATT GTGC
LT-MT: GGGCTTCCCCCTTTcgaa
CGCGGC
CGCGGCCCC
C
kDa
pcDNA3.1
pcDNA3.1-Pou5f1
pcDNA3.1-Sox2
pcDNA3.1-Tcf7l1
IB: α-Oct4
IB: α-Sox2
IB: α-Tcf3
50
40
30
80
60
D
N.E.
Con
O4
S2
T3
LT WT
LT MT
OS WT
OS M1
OS M2
OS M3
Comp
Ab
α-O4
α-S2
α-T3
IgG
*
B
A
E
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
DIG-labeled OS-WT
DIG-labeled LT-WT
E
S2L
S2S
C
D
21 22 23 24 25

Figure 2.10 (Continued)
Physical association of β-catenin, Oct4, Sox2, and Tcf3 and in vitro binding properties of Oct4, Sox2, and Tcf3 to an Oct-Sox composite motif.
(A) Co-immunoprecipitation analysis of β-catenin, Tcf3, Oct4, and Sox2 complexes in mESCs. IP, immunoprecipitation; IB, immunoblotting. Asterisk indicates heavy chains of antibodies used in IP.
(B) Sequence of oligonucleotide probes used in EMSA. Oct motif is underlined; Sox and Lef/Tcf motifs are bolded. Mutations are shown in lowercases.
(C) Expression of Oct4, Sox2, and Tcf3 in 293T cells over-expressing each of the three factors for EMSA. Forty-eight hours after transfection of the indicated plasmids, nuclear extracts were isolated from the cells. Protein expression was determined by western blotting and subsequent immunoblotting using specific antibodies against each factor. IB, immunoblot.
(D) Binding of Oct4 and Sox2 alone to the OS composite motif and Tcf3 to the LT motif.
(E) Cooperative bindings of Oct4, Sox2 and Tcf3 to OS composite motif. Probe sequences are indicated in Figure 2.10B. Bands A, B, C, and D denoted with arrows indicate Oct4-binary, Sox2-binary, Oct4-Sox2-ternary, and Oct4-Tcf3-ternary complexes with the Oct/Sox probe, respectively. Band E indicates Tcf3-binary complex with the LT probe. Asterisks indicate non-specific bands. NE, nuclear extracts; Con, extracts from mock-transfected cells; O4, extracts from Oct4-overexpressing cells; S2, extracts from Sox2-overexpressing cells; S2S, smaller amount of extracts (0.2 mg) from Sox2-overexpressing cells; S2L, larger amount of extracts (2 mg) from Sox2-overexpressing cells. T3, extracts from Tcf3-overexpressing cells; Comp, unlabeled competitors; Ab, antibodies; a-O4, anti-Oct4 antibody; a-S2, anti-Sox2 antibody; a-T3, anti-Tcf3 antibody.

Several lines of evidence support this view. First, the additional band was eliminated by unlabeled LT probe or anti-Tcf3 antibodies (Figure 2.10E, lanes 19 and 22), but not by unlabeled mutated LT probe (Figure 2.10E, lane 20). Second, the additional band was competed by unlabeled WT Oct/Sox probe, by a probe with the Oct-motif mutated, but not by a probe containing mutations in both Oct and Sox motifs (Figure 2.10E, lanes 15, 16, and 18). Finally, when Oct4 binding was competed by unlabeled Sox-mutated probe, or blocked with anti-Oct4 antibodies, the additional band disappeared (Figure 2.10E, lane 17 and lane 21, respectively). Together, these results suggest that Tcf3 binds to the Sox site in the Oct/Sox composite motif in an Oct4-dependent manner whereas Oct and Sox factors can independently associate with their target sites.

To clarify possible patterns of complex formation when all of the three proteins were present we incubated the Oct-Sox composite motif with Oct4, Sox2, and Tcf3 (Figure 2.10E, lanes 23-25). When low amounts of Sox2 were present with the other two proteins, we observed band shifts indicative of Oct4-DNA, Sox2-DNA, and Oct4-Tcf3-DNA complexes (Figure 2.10E, lanes 24; bands A, B, and D, respectively). With higher concentrations of Sox2 we observed the formation of an additional Oct4-Sox2-DNA complex (Figure 2.10E, lane 25, band C): a competition between Sox2 and Tcf3 at target sites has been computationally predicted (Mason et al., 2010). Together these data suggest a mutually exclusive competitive interaction for Tcf3 and Sox2 at Oct-Sox motifs where the association of Tcf3 requires a cooperative interaction with Oct factors to overcome the less favored consensus of Sox versus Lef/Tcf binding motifs.

The functional significance of canonical Wnt signal directed complexes at an Oct-Sox motif was supported by *in vitro* luciferase reporter assay using the *Pou5f1* distal enhancer region (Figure 2.11A). The region belongs to Group B in Figure 2.5B-G: the sequence of the Oct/Sox

probe used in the EMSA analysis above derives from this region. We confirmed copy number-dependent enhancer activity specific to 2i-cultured mESCs (v6.5), not observed in NIH3T3 cells, (Figure 2.11B). Further, 2i-cultured mESCs showed an up-regulation of enhancer activity compared to PD03-cultures, suggesting a positive effect of CHIR stimulus, and consequently, canonical Wnt input into enhancer activity; up-regulation was notably diminished upon mutation of the Sox motif to the similar extent as mutation of both Oct and Sox motifs. In contrast, when the Oct motif was mutated, we still observed elevated enhancer activity, though to a lesser extent compared to wild type, when 2i-culture was compared to PD03-treated alone (Figure 2.11C, left). The similar trend was recapitulated in mESCs cultured with serum+LIF (Figure 2.11C, right). Together, these data support the conjecture that canonical Wnt signaling contributes to the transcription of pluripotency genes via the Sox site within an Oct-Sox composite motif.

Figure 2.11

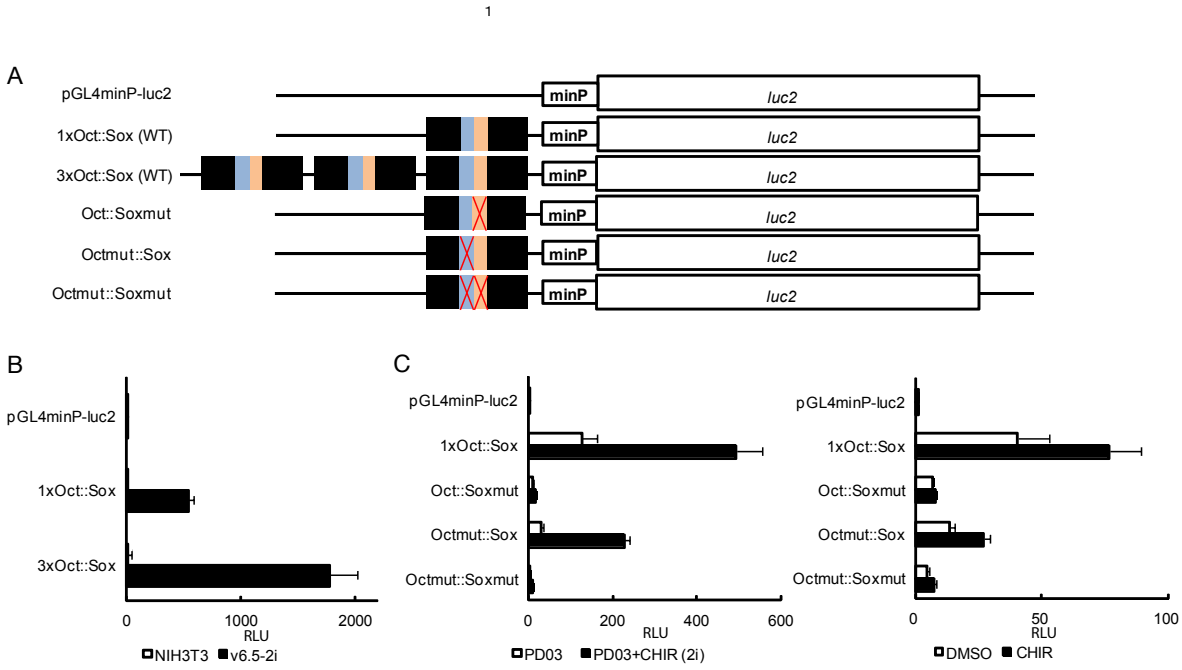

1
A
pGL4minP-luc2
1xOct::Sox (WT)
3xOct::Sox (WT)
Oct::Soxmut
Octmut::Sox
Octmut::Soxmut
minP
luc2
B
pGL4minP-luc2
1xOct::Sox
3xOct::Sox
0
1000
2000
RLU
NIH3T3
v6.5-2i
C
pGL4minP-luc2
1xOct::Sox
Oct::Soxmut
Octmut::Sox
Octmut::Soxmut
0
200
400
600
RLU
PD03
PD03+CHIR (2i)
pGL4minP-luc2
1xOct::Sox
Oct::Soxmut
Octmut::Sox
Octmut::Soxmut
0
50
100
RLU
DMSO
CHIR

Figure 2.11 (Continued)
***In vitro* luciferase assay shows that canonical Wnt signaling contributes to the transcription of pluripotency genes via the Sox site within an Oct-Sox composite motif.**
(A) Schematic of luciferase reporter constructs used in (B) and (C). Pou5f1 distal enhancer region containing the Oct-Sox composite motif drives the luciferase gene with a minimal TATA-box promoter element under pGL4 vector backbone. Each mutation corresponds to mutant motifs in EMSA analysis.
(B) Luciferase reporter assay using Pou5f1 distal enhancer region in NIH3T3 cells and 2i-cultured mESCs (v6.5). Forty-eight hours after transfection with reporter constructs, cells were subjected to the assay. mESCs were maintained under 2i+LIF condition for 11 passages prior to the assay.
(C) Luciferase reporter assay in mESCs (v6.5) cultured under 2i condition (left) or serum+LIF (right). mESCs were maintained under 2i+LIF condition for 11 passages prior to the assay. Upon transfection with reporter constructs, cells were switched into basal media of 2i culture (mixture of neurobasal media, DMEM/F12, N2, and B27 supplements) with PD03 or 2i (left), or CM in the presence or absence of CHIR (right). The assay was performed 48 hours after transfection. RLU, relative light unit.

Discussion

Our transcriptional analysis of Wnt pathway action in ESCs has generated several new insights into pluripotency and differentiation networks. First, ChIP-seq analysis enabled the identification of genomic targets of β-catenin activity in mESCs. Second, a strong association is observed between β-catenin bound regions and those occupied by core pluripotency factors (NOS) and Tcf3. Third, there are marked differences in bound regions and candidate target genes between regions where only β-catenin::Tcf3 overlap, and those where β-catenin::Tcf3 also intersect with the core pluripotency networks (NOS). This is observed in motif enrichment (Lef/Tcf motif vs. Oct-Sox motif) suggesting distinct binding modes, the function of associated genes (axis specification- and mesoderm-related genes vs. stem cell- and ectoderm-related genes) suggestive of different biological outcomes, and the activity status of likely enhancer regions in mESCs (high activity for β-catenin::Tcf3::NOS and low activity for β-catenin::Tcf3). Fourth, under standard culture condition, the activation of canonical Wnt signaling elevated expression of differentiation-related genes, while activity of pluripotency-related genes was maintained. Fifth, under 2i condition, β-catenin also engaged at likely enhancers for TE lineage- and axis specification-related genes but under 2i conditions, these targets are not activated. Inhibition of MEK/ERK signaling by PD03 is critical in blocking these differentiation pathways and the enrichment of Ets motifs within differentiation related enhancers suggest a cooperative interplay of Ets and canonical Wnt complexes for gene activity. Sixth, β-catenin, Tcf3, and Oct4 interact under 2i conditions. Finally, canonical Wnt signaling up-regulated transcription in cell culture utilizing Oct-Sox motifs that could also be engaged directly by Oct4-Tcf3 and Oct4-Sox2. Considering all data, we propose that under the 2i condition, canonical Wnt signaling participates in the pluripotency network via Oct4/β-catenin/Tcf3 complex formation and

although Sox2 is absent, engagement of β-catenin still favors activity of pluripotency-associated genes (Figure 2.12, upper).

The analysis of canonical Wnt pathway mutants sheds additional light on this process. First, β-catenin activity is essential under 2i conditions for the derivation and maintenance of mESCs (Wray et al., 2011). Second, our work, and that of others indicate that Tcf proteins are required for occupancy at Oct4-dependent promoter (Kelly et al., 2011; Wray et al., 2011; Yi et al., 2011). Further, engagement of β-catenin can enhance Oct4 promoter activity (Kelly et al., 2011; Wray et al., 2011; Yi et al., 2011). Tcf3 is a key component in the transcriptional complex; Tcf3 actions are linked to Groucho-mediated gene silencing (Brantjes et al., 2001). Tcf3 inhibits pluripotency and the removal of Tcf3 can functionally substitute for the actions of β-catenin under 2i conditions (Guo et al., 2011; Schaniel et al., 2009). Thus, the main action of β-catenin appears to be to neutralize the destabilizing activity of Tcf3 in the pluripotency network. Interestingly, mutant forms of β-catenin lacking the transcriptional activating domain are effective in maintaining pluripotency (Wray et al., 2011) suggesting that β-catenin acts by abrogating Tcf3 silencing rather than forming a β-catenin-dependent activation complex. The conclusion that Tcf3 and β-catenin do not form an active transcriptional complex is supported by recent studies of Tcf3 mutant mouse embryos (Wu et al., 2012).

Wnt actions in maintaining a state of pluripotency have also been linked to the control of telomerase activity through direct regulation of *Tert* promoter activity in mESCs. β-catenin binding was reported to be enriched around the *Tert* gene in ChIP analysis of ESCs and binding further enhanced by Wnt3a treatment, or expression of a stabilized form of β-catenin (Hoffmeyer et al., 2012). In contrast, we see no enrichment at the *Tert* locus in our whole genome analysis (data not shown).

Figure 2.12

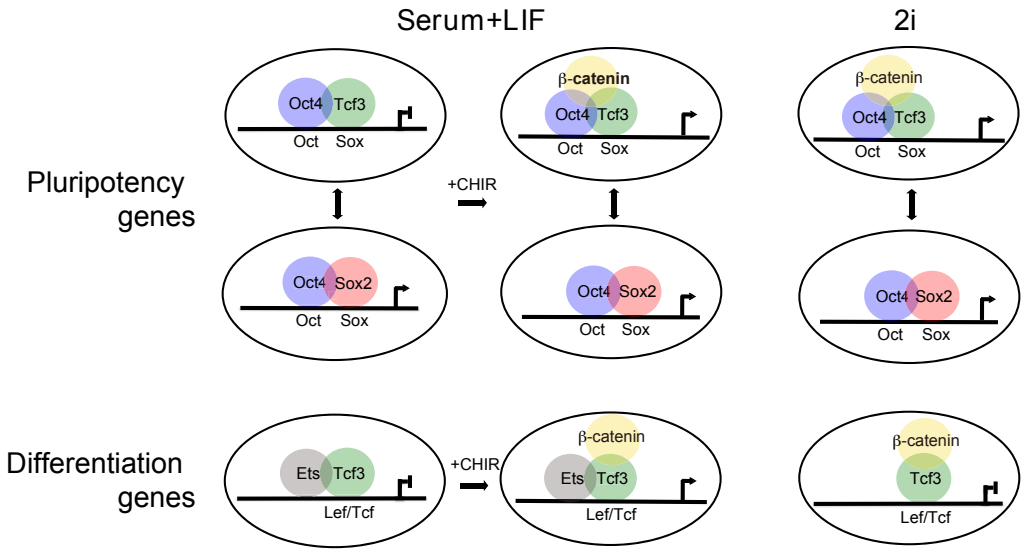

Serum+LIF
2i
Pluripotency genes
Differentiation genes
+CHIR
+CHIR
β-catenin
Oct4
Tcf3
Sox2
Oct
Sox
Ets
Lef/Tcf

Figure 2.12 (Continued)
Schematic model of β-catenin-dependent regulation of pluripotency network.
Oct4-Sox2 binding to Oct-Sox composite motifs maintains activity of key regulators of pluripotency. Tcf3 interaction with Oct factors at the same motif is predicted to destabilize this circuit. CHIR-mediated stabilization of β-catenin has opposing actions. Entry of β-catenin into Oct4/Tcf3 complexes abrogates Tcf3 actions thereby promoting pluripotency. However, the production of active canonical Wnt transcriptional complexes engages differentiation targets destabilizing pluripotency. PD03-mediated inhibition of MEK/ERK signaling restores a pluripotency balance blocking the activation of Wnt dependent differentiation genes enabling culture under 2i conditions. Given the role of MEK/ERK signaling downstream of receptor tyrosine kinases in the regulation of the Ets-family of transcriptional regulators, and the enrichment of Ets motifs in predicted *cis*-regulatory, we propose the combined action of Wnt and RTK signaling in the differentiation of ES cells.

Summary

ESCs are in an inherently unstable state wherein endogenous Tcf3 activity antagonizes the core pluripotency network through competition at Oct/Sox motifs for Sox2 binding. Surprisingly, analysis of Sox2, Nanog, and Oct4 binding suggest that Tcf3 is not a transcriptional target of this network (data not shown). Under standard culture conditions of serum and LIF, β-catenin plays no significant role in maintaining pluripotency: its activity is not essential (Lyashenko et al., 2011) (Wray et al., 2011) and stimulation of the pathway favors differentiation through engagement at "classic" Lef/Tcf motifs around differentiation associated target genes. Presumably, LIF actions are dominant and the levels of activation of target genes are not sufficient to trigger widespread differentiation. In contrast, in 2i medium Tcf3 actions are critical and β-catenin is essential to overcome Tcf3's inhibitory effects within the pluripotency network. What remains to be determined is how the inhibition of MEK/ERK signaling blocks the differentiation promoting arm of Lef/Tcf::β-catenin -directed gene regulation. Our data on motif recovery and β-catenin engagement suggest a cooperative role for MEK/ERK directed Ets factors independent of Lef/Tcf::β-catenin binding to activate target differentiation promoting genes consistent with the critical actions of Fgf and Wnt signaling in promoting lineage commitment in early mammalian development (Figure 2.12, lower).

Materials and methods

Generation of ES cell lines

For the construction of *Ctnnb1* targeting vectors, homologous arms were sub-cloned by PCR from a BAC clone RP24-499-N24 using Platinum *pfx* DNA Polymerase (Invitrogen, 11708-013). 5' homologous arm (chr9:120,864,661-120,868,632) carrying *Aat II* and *Nde I* sites at its C-

terminus was cloned into *Sac II* and *Not I* sites of the pPGKneobpAlox2PGKDTA (generously provided by P. Soriano, Mount Sinai School of Medicine, NY, U.S.A), upstream of the PGKneo cassette flanked by two loxP sites. *Biotin-3xFLAG* sequence (5'-GGGTCCGGCCTGAACGACAT
CTTCGAGGCTCAGAAAATCGAATGGCACGAAGGCGCGCCGAGCTCGAGG-
GACTACAAAGA
CCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAG-3') was cloned into the *Aat II* and *Nde I* sites of 5' homologous arm with stop codon. 3' homologous arm (chr9:120,868,654-120,871,607) was cloned into the *Hind III* site of the pPGKneobpAlox2PGKDTA, downstream of the PGKneo cassette. The final targeting vector was linearized with *Apa I* and electroporated into 129/Sv x C57BL/6J F1 hybrid ESC (v6.5). After the expansion of G418-resistant clones, homologous recombination was screened by PCR using 5' and 3' external primers on genomic DNA (Fig S1B). Primer sequences used in the screening are as follows: P1, 5'-ttgtctcagaatagaacagaggaatgttac-3' (in 5' external region of the homologous arm); P2, 5'-atccttgtagtcgatgtcatgatctttat-3' (within *Biotin-3xFLAG* sequence) ; P3, 5'-gattataaagatcatgacatcgactacaag-3' (within *Biotin-3xFLAG* sequence); P4, 5'-ttggtcgaggagtaacaatacaaat -3' (in the 3' homologous arm); P5, 5'-atgtgactagcagctttaaccaagag-3' (in 3' external region of the homologous arm). A clone in which homologous recombination was confirmed by the PCR (2A3) was used as the *Ctnnb1-BioFLneo* ESC.

To generate BirA-expressing *Ctnnb1-BioFLneo* ESC lines (*Ctnnb1-BioFLneo;BirA* ESC) and v6.5 (*BirA* ESC), linearized pCAGGS-BirA-V5-IRES-Hyg was electroporated into *Ctnnb1-BioFLneo* ESC clone 2A3 or v6.5 cells. After the expansion of Hygromycin-resistant clones, BirA-expressing clones were screened by immnoblotting for the V5 epitope. For the construction

of the pCAGGS-BirA-V5-His-IRES-Hyg, we first subcloned IRES-Hyg cassette, of which internal *Not I* site was mutated, from pIRES-Hyg3 vector (Clonetech, 631620). The IRES-Hyg cassette was exchanged with IRES-dsRed on pCIR (Tenzen et al., 2006) using *Sma I* and *Not I* sites to generate the pCAGGS-IRES-Hyg vector. BirA-V5-His fragments derived from the pEF1a-BirA (generously provided by A. B. Cantor, Children's Hospital Boston, MA, U.S.A) were then cloned into the *Eco RV* site of pCAGGS-IRES-Hyg vectors to make the pCAGGS-BirA-V5-IRES-Hyg.

ESC culture

All ESCs were maintained in complete ES media (15% fetal bovine serum, 0.1 mM non-essential amino acids, 0.1 mM β-mercaptoethanol, 2 mM L-glutamine, 10^3 units/ml LIF, and 1X nucleotide mix in DMEM high glucose) with feeder cells isolated from DR4 mice (Tucker et al., 1997). 2i culture of ESCs was performed as described using CHIR99021 (3 mM, Stemgent, 04-0004) and PD0325901 (1 mM; Stemgent, 04-0006) (Ying et al., 2008a).

Immunofluorescence, Immunoblot, and Co-immunoprecipitation Assays

Cells cultured on gelatinized chamber slides were fixed with 4% paraformaldehyde/PBS. After blocked with 0.1% Tween, 3% BSA (Sigma, A7906), and 1% heat inactivated sheep serum (Sigma, S2263) in PBS, slides were incubated with anti-FLAG M2 (1:500; Sigma, F3165) and anti-β-catenin antibodies (1:500; Epitomics, 1247-1).

Whole cell lysates were obtained using RIPA buffer (10 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, and 1% deoxycholic acid; pH 7.8). Nuclear extracts were isolated according to the manufacturer's instructions for Nuclear Complex Co-IP Kit (Active

Motif, 54001). Co-immunoprecipitation was performed according to the manufacturer's instructions for Protein A/G HP SpinTrap Buffer Kit (GE Healthcare, 28-9135-67). Lysates or immunoprecipitated materials were subjected to SDS-PAGE and blotted onto nitrocellulose membranes. The following antibodies were incubated overnight at 4°C in PBS containing 3% non-fat milk and 0.1% Tween20: HRP-conjugated anti-FLAG M2 antibody (1:500; Sigma, A8592), HRP-conjugated streptavidin (1:500; Perkin Erlmer, NEL750), anti-V5-HRP antibody (1:5000; Invitrogen, R961-25), anti-β-catenin (1:1000; Epitomics, 1247-1), anti-Tcf3 (1:200; Santa Cruz Biotechnology), anti-Oct4 (1:200; Santa Cruz Biotechnology), and anti-Sox2 (1:400; R&D Systems) antibodies. HRP-conjugated secondary antibodies for western blotting were purchased from Southern Biotech and Jackson Immunoresearch.

ChIP, ChIP-seq and ChIP-qPCR

After treated with CHIR99021 (3 mM) for 16 hours, *Ctnnb1-BioFLneo;BirA* and *BirA* ESC were collected with feeder cells using 0.25% trypsin-EDTA. After leaving cells on gelatinized plates for an hour, we collected floating cells as an ESC-enriched fraction. 10,000,000 cells were used per ChIP. Chromatin was sonicated by 6 sessions of 30 pulses (1 sec on and 1 sec off) at 50% amplitude using the Branson sonifier 250D. Isolation of chromatin and FLAG-ChIP were performed as described (Vokes et al., 2007). For Biotin-ChIP, sonicated chromatins were first incubated with protein A dynabeads (Invitrogen, 100-01D) at 4 °C for an hour. Dynabeads MyOne Streptavidin T1 (Invitrogen, 656.01) was incubated with 2% Gelatin from cold water fish skin (Sigma, G7041) in PBS at 4 °C for an hour prior to use. Pre-cleared chromatins were incubated with 60 mL of Dynabeads MyOne Streptavidin T1 overnight with rotation at 4 °C. Beads were washed with 2% SDS, 10 mM Tris, and 0.5 M EDTA five times, and with TE once.

ChIP DNA was eluted by the overnight incubation of beads in 10 mM Tris, 1 mM EDTA, 1% SDS, and 0.5 M NaCl at 65 °C. DNA was purified as described (Vokes et al., 2007). ChIP DNA was analyzed by SYBR green real-time PCR (see primers in Table 2.1). Enrichment was calculated relative to the control region. ChIP-seq libraries were constructed using ChIP-seq DNA Sample Prep Kit (Illumina, IP-102-1001) according to manufacturer's instruction. We also generated input control library for both anti-FLAG and Biotin replicates, in addition to the mock IP controls generated from BirA ESC.

For ChIP in 2i culture system, 2i-adapted mESCs cultured in 2i+LIF were treated with PD0325901 (1 μM), CHIR99021 (3 μM), or 2i (PD0325901+CHIR99021) for 24 hours and subjected to ChIP procedures.

ChIP-qPCR was performed with Biorad iQ™ SYBR® Green Supermix (#170-8880). Fold enrichment was calculated by normalizing ChIP sample against input, and target region against control region as follows.

$\Delta Ct = Ct(ChIP) - Ct(input)$;

$\Delta\Delta Ct = \Delta Ct(target\ region) - \Delta Ct(control\ region)$;

$Fold\ enrichment = 2^{-\Delta\Delta Ct}$

Expression Profiling

For culture in CM with feeder cells, *Ctnnb1-BioFLneo;BirA* ESCs were transferred onto gelatinized plates 24 hours before the 24-hour-treatment of CHIR99021 (3 μM) or XAV939 (1 μM; Stemgent, 04-0046) (Hall et al., 2001; Huang et al., 2009b; Ring et al., 2003); for culture in 2i system, *Ctnnb1-BioFLneo;BirA* ESCs were treated with DMSO only, PD0325901 (1 μM) only, CHIR99021 (3 μM) only, 2i (PD0325901 1μM +CHIR99021 3μM) or 2i+LIF for 24 hours

After the treatments, total RNA was isolated using TRIZOL Reagent (Invitrogen, 15596-026). RNA was then treated with DNaseI (QIAGEN, 79254) and purified by RNeasy Mini Kit (QIAGEN, 74104) according to manufacturers' instruction. 300 ng of total RNA were labeled for each sample using the GeneChip WT cDNA Synthesis and Amplification Kit and WT Terminal Labeling Kit (Affymetrix, 900673 and 900671). Labeled sense-strand DNA targets were hybridized onto GeneChip mouse Gene 1.0 ST array (Affymetrix, 901169). Hybridization signal was scanned on the GeneChip Scanner 3000 (Affymetrix). All array-related procedures were performed according to Affymetrix GeneChip protocols.

Data analysis was performed by loading CEL files into the statistical computing language R software, and probes were normalized using robust multi-array average (RMA) (Irizarry et al., 2003). LIMMA was used for differential gene expression analysis, and adjusted P-value < 0.005 were considered differentially expressed, unless otherwise noted.

Expression microarray of ESCs, mesoendoderm cells, and neuroectoderm cells were from GEO GSE12982 (Shen et al., 2008). Correlation plots of Figure 2.7B were generated in the following way. Top differentially expressed genes in MEC/ESC and NEC/ESC in different tiers (step size: 20 genes) were extracted. The Pearson correlation of their fold changes in ME/ES or NE/ES with that in CM+CHIR/CM+XAV were calculated using R.

EMSA

Pou5f1 and *Sox2* cDNA were provided by Dr. Taku Saito (The University of Tokyo Graduate School of Medicine, Tokyo, Japan); *Tcf7l1* cDNA was obtained from Open Biosytems. Plasmids expressing *Pou5f1*, *Sox2*, or *Tcf7l1* were constructed by cloning each cDNA into pcDNA3.1(-) vectors (Invitrogen, V795-20). 293T cells were transfected with pcDNA3.1(-)-*Pou5f1*,

pcDNA3.1(-)-*Sox2*, pcDNA3.1(-)-*Tcf7l1*, or pcDNA3.1(-) empty vectors using FuGENE 6 Transfection Reagent (Promega, E2691). Forty-eight hours after transfection, nuclear extracts were isolated as described (Dignam et al., 1983). Isolated extracts were then dialyzed with Slide-A-Lyzer G2 Dialysis Cassette (Pierce Biotechnology, 87717) in dialysis buffer (20 mM HEPES, pH 7.9, 20% glycerol, 100 mM KCl, 0.83 mM EDTA, 1.66 mM DTT) supplemented with protease inhibitor cocktail (Complete, EDTA-free; Roche, 1873580) for 1 hour at 4 °C. The concentration of nuclear extracts was determined by the Qubit fluorometer and Quant-iT Protein Assay Kit (Invitrogen, Q33211).

Probes were labeled using DIG Gel Shift Kit, 2nd generation (Roche, 3353591) according to manufacturer's instruction. Two microgram of nuclear extracts were added to the binding mixture containing 12 mM HEPES (pH 7.9), 12% glycerol, 60 mM KCl, 0.5 mM EDTA, 1 mM DTT, 1 mg of poly (dG-dC)·poly (dG-dC) (Sigma, P9389), and 30 fmol of the digoxigenin (DIG)-labeled probe. Binding reaction was performed for 15 min at room temperature. Competition assay was performed by incubating nuclear extracts with 133-fold molar excess of unlabeled probes for 30 minutes on ice prior to binding reaction. When antibodies were used, nuclear extracts were incubated with 4 mg of anti-Oct-3/4 (sc-9081x, Santa Cruz Biotechnology, Inc.), 6 mg of anti-Sox2 (sc-17320x, Santa Cruz Biotechnology, Inc.), or 6 mg of anti-TCF3 (sc-8635x, Santa Cruz Biotechnology, Inc.) for 30 minutes on ice prior to binding reaction. Oligonucleotide probes used are as follows: wild-type Oct/Sox motif-containing probe (OS-WT), 5'-GGCAGATGCATAACAAAGGTGC-3'; OS mutant 1 (OS-M1), 5'-GGCAGCCCCATAACAAAGGTGC-3'; OS mutant 2 (OS-M2), 5'-GGCAGATGCATAACACCCGTGC-3'; OS mutant 3 (OS-M3), 5'-

GGCAGCCCCATAACACCCGTGC-3'; LRE-WT, 5'-TTCCCCCTTTGATTCGCGGC-3'; LRE mutant (LRE-MT), 5'-GGGCTTCCCCCTTTCGAACGCGGCCCC-3'.

Binding mixtures were run on pre-run 6% TBE gels (Invitrogen, EC62652BOX) in 0.25x tris-borate-EDTA (TBE) buffer at 150 V for 60 min. Gels were transferred onto positive-charged nylon membranes (Roche, 1209272) by electro-blotting at 400 mA for 30 minutes in 0.5x TBE buffer. After cross-linked at 120 mJ with a transilluminator, DIG-labeled probes were detected with anti-Digoxigenin-AP (Roche, 1093274) in combination with CDP-star (Roche, T2145) according to manufacturer's instruction. Images were taken with ImageQuant LAS 4000mini (GE Healthcare UK Ltd.).

Luciferase reporter assay

A 117 bp fragment corresponding to the β-catenin peak region 2.1 kb upstream of mouse *Pou5f1* gene was obtained by PCR using Platinum pfx DNA polymerase (Invitrogen) and cloned into pBluescript II SK (-) (Stratagene, 212206) at Kpn I and Hind III sites (pBS-Oct::Sox). In the PCR, Kpn I-Sal I and Hind III-Xho I sequences were added to 5' and 3' ends of the fragment, respectively. Cloning of tandem fragments were performed using Sal I and Xho I sites on the pBS-Oct::Sox. The single or triplet copies of fragments were then transferred into pGL4.26[*luc2*/*minP*/*Hygro*] vector (Promega, E844A) at Kpn I and Hind III sites.

For luciferase reporter assays, cells were plated on 24-well plates and transfected with 0.4 mg of pGL4 reporter constructs and 8 ng of pGL4.74[*hRluc*/*TK*] using FuGENE 6 transfection reagent (Promega, E2691). pGL4.74[*hRluc*/*TK*] was used for the normalization of the transfection efficiency. Forty-eight hours after transfection, luciferase activities were measured

using Dual-Luciferase Reporter Assay System (Promega, E1910) and GloMax 96 Microplate Luminometer (Promega).

ChIP-seq data analysis

The first 25 bp of raw reads were used for mapping against mouse mm9 genome assembly (NCBI Build 37). If a read has a unique mapping location (2-mismatch allowed) in the genome, we call it a mappable read. If in this sequencing data set, only one read maps to this genomic location, we call it a non-redundant read. In peak calling, we compared using input control versus mock IP control, and no significant differences were observed. Peak detection was performed using MACS software (Zhang et al., 2008) with a p-value cutoff of 1e-5 and FDR<0.05 using input controls. Identified peaks were normalized to 400 bp each by extending 200 bp upstream and downstream of peak summit. Enrichment peaks were annotated using CisGenome package build 37 of the mouse genome, and CisGenome Browser was used to visualize the peak signal (Ji et al., 2008). Other published sets of ChIP-seq data were downloaded from Gene Expression Omnibus (Chen et al., 2008; Creyghton et al., 2010; Han et al., 2010; Heng et al., 2010; Ku et al., 2008; Marson et al., 2008b; Meissner et al., 2008; Mikkelsen et al., 2007). Raw data were processed in the same way as the β-catenin ChIP-seq data. Over-represented motif analysis were performed using Cistrome Analysis Pipeline (Liu et al., 2011), and motif positional weight matrix mapping to genome coordinates were performed by CisGenome package using the default setting of likelihood ratio 500 (Ji et al., 2008). The distribution of motifs near peak summit and motif logos were plotted in R. For the matched control regions, for each peak in the input, we calculated its distance to the nearest TSS, and randomly chose a gene and exacted one matched control region around this gene that had the

same distance to that TSS. The p-value for motif enrichment is calculated by a two-proportion z-test.

Sequencing depth in Figure 2.3B was evaluated by the method modified from Cao et al (Cao et al., 2010). We randomly sampled 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of total raw reads, and calculated the proportion of background reads (defined as the reads that were not within the peaks called under that sequencing depth) that were converted to foreground reads (the reads that were within the peaks called under that sequencing depth). If sequencing depth is sufficient, we expect that the rate for this conversion will decrease to an asymptote of zero in the ideal case. Both replicates reach plateau at around 80% total sequencing reads, indicating enough reads have been sequenced to have sufficient coverage.

Gene and enhancer annotation

Gene ontology analyses were performed using DAVID (http://david.abcc.ncifcrf.gov/) (Huang et al., 2009a). The association of genome regions with gene ontology terms was done using GREAT (version 1.8.2) (http://great.stanford.edu/) (McLean et al., 2010).

Histone modification analysis

Aggregates plot of histone modifications were performed using Homer (Heinz et al., 2010). Genomic regions 3 kb upstream and 3 kb downstream from the peak summit were divided into 60 100-bp bins each. Error bars indicate standard errors of the means of all the normalized ChIP-seq fragment densities in the respective bin. Box plot of expression values was drawn in R. Statistical significance was calculated by the Wilcoxon rank-sum test.

Direct target gene prediction

Direct target genes of β-catenin were predicted by the method described by Tang et al (Tang et al., 2011). The FDR was calculated based on a permutation method described in (Breitling et al., 2004). The scatter plots in Figure 2.7A was generated using R.

Hierarchical clustering, heat map and Venn diagram

For the hierarchical clustering of 20 factors based on chromatin co-occupancy, peak regions for each factor with p-value < 1e-8 were trimmed or expanded to 400 bp centered at the peak summit. Pair-wise intersection was performed between every pair of factors to generate a matrix of the number of co-bound regions. Complete linkage, correlation distance, hierarchical clustering was performed and heat map generated using R. Peak-peak intersection were conducted using BEDTools (v2.10.1) (Quinlan and Hall, 2010). Venn diagrams were draw in Cistrome analysis pipeline(Liu et al., 2011).

Kohonen Self-organizing map (SOM)

Kohonen SOM was generated in Matlab using SOM Toolbox (http://www.cis.hut.fi/somtoolbox/). Genes with expression fold change > 2 in at least one comparison group of 2i/PD03, 2i/CHIR, 2i/DMSO, 2iLIF/2i, 2iLIF/DMSO, and CM+CHIR/CM+XAV were extracted and subjected to K-means clustering (Figure 2.8D), individual component analysis (Figure 2.8E), and U-matrix generation (Figure 2.9).

Table 2.1

Table-2.1: Sequences for ChIP-qPCR primers and genomic coordinates for peaks tested.

Loci name	Peak Genomic Coordinates			Primer-Forward	Primer-Reverse	Position
T-Up0.1k	chr17	8626866	8627266	CTTTGATGGAGGTGCAAACAT	ACCCTTTGAAGTACCGAGCAG	Upstream
Porcn-Up13k	chrX	7796887	7797287	CCTTATCACTGAGAGGAGGAACA	AATAACTTGTGGGCAAAGGAGAT	Upstream
Cdx4-In5.3k	chrX	100521894	100522294	TCTGACATTTGTTTAGGGCTTGT	TAAAGTGGGTCCCTAAACTCCAT	Intron
Eomes-Up9.8pk1	chr9	118377272	118377672	GAGTGTTCCATCTCTCAAAAGGA	GGTAAACAAGGTCTGGGACTAGG	Upstream
Eomes-Up9.8pk2	chr9	118380437	118380837	GCTCCCTAGAGAGGTCTCAGAAT	AGGAGCTGAGGCTAAGTGAAACT	Upstream
Cdx2-In2.8k	chr5	148115783	148116183	AAAACCTTTGCAAACCACTACAA	GGGCTGTGTTTATGTTAATGCTC	Intron
Klf4-Down57k	chr4	55488284	55488684	GAATGAAGTTGTGTGGAGGTCAT	ATTCCTAGAAGTGGCCTCAAAAG	Downstream
Klf4-Down68k	chr4	55477019	55477419	CAGTAAACCATTGTTGTTGACCA	TTAAAGGGTACGGAAGAATAGGG	Downstream
Klf2-Up2.9k	chr8	74834952	74835352	GTCTGAGTCAGCTCTCCAAATGT	TCACAGACCTCTCTTGAGACCTT	Upstream
Klf2-Down87k	chr8	74858115	74858515	CTTGGGAAGAATCTGCACTTAAC	GGCATTTCAAAGTCACACTCTTT	Downstream
Epha1-Up7.7k	chr6	42330763	42331163	AGAAAAGGAGTTTTCTGCGAAGT	GAGGAAGAAAAGGAACGCTACAT	Upstream
Tead1-In12k	chr7	119834629	119835029	CATTTGTGATTGAGACCTGTGAA	CTCTTTGTTTCCACATCAAGGAC	Intron
Tead1-In9k	chr7	119832052	119832452	TTTAAGGGGACAGGACTTTTGTT	AACAGCTGACTTCCAAAACTACG	Intron
Lin28-Up2k	chr4	133576534	133576934	GGTCTTCTGAAGGGAAGAAGAAA	GAAAGATGCAAAGCAGGAAAGTA	Upstream
Tcfcp2l1-Up2.3k	chr1	120522039	120522439	GGTTCATCTGCATATCAGTCCTC	AAGTGGATTCTTTTGTCCAAGTG	Upstream
Nanog-Up4.7k	chr6	122652653	122653053	GCTTCCCTGGATAAGGAATGTAG	CCACCATAGCCTTAAGTTTACCC	Upstream
Nanog-TSS	chr6	122657229	122657629	TTGACCTGAAACTTCCCACTAGA	GGACATTGTAATGCAAAAGAAGC	TSS
Pou5f1-Up0.9k	chr17	35641866	35642266	TATCTCCATCTGAGGCTCTGTCT	GGCAGCTCTAACCCTAAACAAGT	Upstream
Pou5f1-Up1.9k	chr17	35640797	35641197	CCCAGGGAGGTTGAGAGTTC	TTGTGGAACAGTGCCATAGG	Upstream
Utf1-Down1.8k	chr7	147131387	147131787	TCCTCAGGGACTAGAGAGTCAGA	GGACTTCCCTTAGCCAAGACTTA	Downstream
Lef1-Up4.4k	chr3	130808751	130809151	TCCAGGCTTGATTTTTATTTTCA	TGTGTGACTAAATTTGGCAAGAA	Upstream
Lef1-Up8k	chr3	130805157	130805557	GAGGAAGGCCTGAAGTATCTACC	AATAAAGCCCTTTGTTTCATGCT	Upstream
Msx2-Down11.2k	chr13	53556747	53557147	TAAAGAAAGTAGGGAGGCTTGCT	AGAAATCCACACAGCTCAAGAAG	Downstream
Sp5 (TSS)	chr2	70312855	70313150	CCTCAGTGTGAGGATGCAGA	GGCAGTGCTCAAAGTGACAA	TSS
Axin2 (TSS)	chr11	108466633	108466948	GTGCGCCAGCGGATCAATGGTGAGT	AATAGCCGGCCTGCCAACTTCAAAG	TSS
Id3-Down3.1k	chr4	135702721	135702964	AGTGGAGCAAGCCGCTTCTCTCTG	CTCAGAGGTTTCTGACCTAGGGCTAAAG	Downstream
Tbx3-In9.3k	chr5	120129896	120130187	AGTACTCCAGCAGAGTTCTGAGGTCACG	TCCCAGACTTCATTTTCCCAGTCTG	Intron
Sox2-Up20k	chr3	34567416	34567728	CACTCCTAGACTGTTCCCCTCC	ACATATTGAGCTATCTCTCCTGGGAAG	Upstream
Rif1-Up0.5k	chr2	51927635	51927912	TGGGGTCCAATGGAAGTAAA	TAACAAAGGCCTGGAAATGG	Upstream
control*	chr2	121973115	121973242	TCAGGTCCTAAGTCCTTTTCTGAGTGG	TGGACCCTTGGTGCCCTACTATCTAG	

*Nature. 2006 Nov 16;444(7117):364-8.

Acknowledgements

We thank Drs. Philippe Soriano, Alan B. Cantor, and Taku Saito for provision of experimental materials; Laurie Chen, Joe Vaughan, Jill McMahon, Christian Daly, Jennifer Couget, Qianzi Tang, and Genome Modification Facility, Harvard Stem Cell Institute for providing technical assistance. We are also grateful to Dr. Ung-il Chung for critical comments and helpful discussions. Work in A.P.M.'s laboratory was supported by a grant from the NIH (DK056246). X.S.L was supported by a grant from the NIH (2R01HG4069). S.O. was supported by Grants-in-Aid for young scientist (#23689079) from the Japan Society for the Promotion of Science (JSPS), Daiwa Securities Health Foundation Grant, and Nakatomi Foundation Research Grant.

Disclosure of potential conflicts of interest

The authors indicate no potential conflicts of interest.

References

Aberle, H., Bauer, A., Stappert, J., Kispert, A., and Kemler, R. (1997). beta-catenin is a target for the ubiquitin-proteasome pathway. Faseb Journal *11*, A1409-A1409.

Anton, R., Kestler, H.A., and Kuhl, M. (2007). beta-Catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells. Febs Lett *581*, 5247-5254.

Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C.J., Mclauchlan, H., Klevernic, I., Arthur, J.S.C., Alessi, D.R., and Cohen, P. (2007). The selectivity of protein kinase inhibitors: a further update. Biochem J *408*, 297-315.

Bakre, M.M., Hoi, A., Mong, J.C.Y., Koh, Y.Y., Wong, K.Y., and Stanton, L.W. (2007). Generation of multipotential mesendodermal progenitors from mouse embryonic stem cells via sustained Wnt pathway activation. J Biol Chem *282*, 31703-31712.

Barrow, J.R., Howell, W.D., Rule, M., Hayashi, S., Thomas, K.R., Capecchi, M.R., and McMahon, A.P. (2007). Wnt3 signaling in the epiblast is required for proper orientation of the anteroposterior axis. Dev Biol *312*, 312-320.

Behringer, R.R., Liu, P.T., Wakamiya, M., Shea, M.J., Albrecht, U., and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. Nat Genet *22*, 361-365.

Birchmeier, W., Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., and Birchmeier, C. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. Journal of Cell Biology *148*, 567-578.

Brantjes, H., Roose, J., van de Wetering, M., and Clevers, H. (2001). All Tcf HMG box transcription factors interact with Groucho-related co-repressors. Nucleic Acids Res *29*, 1410-1419.

Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. Febs Lett *573*, 83-92.

Brook, F.A., and Gardner, R.L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. Proceedings of the National Academy of Sciences *94*, 5709-5712.

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., Macquarrie, K.L., Davison, J., Morgan, M.T., Ruzzo, W.L., *et al*. (2010). Genome-wide MyoD Binding in Skeletal Muscle Cells: A Potential for Broad Cellular Reprogramming. Developmental Cell *18*, 662-674.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W.W., Jiang, J.M., *et al*. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Chew, J.L., Loh, Y.H., Zhang, W.S., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P., *et al*. (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol Cell Biol *25*, 6031-6046.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Gene Dev *22*, 746-755.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al*. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. P Natl Acad Sci USA *107*, 21931-21936.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc Natl Acad Sci USA *100*, 7480-7485.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate Transcription Initiation by Rna Polymerase-Ii in a Soluble Extract from Isolated Mammalian Nuclei. Nucleic Acids Res *11*, 1475-1489.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Gonzalez, F., Boue, S., and Belmonte, J.C.I. (2011). Methods for making induced pluripotent stem cells: reprogramming a la carte. Nature Reviews Genetics *12*, 231-242.

Guo, G., Huang, Y., Humphreys, P., Wang, X.Z., and Smith, A. (2011). A PiggyBac-Based Recessive Screening Method to Identify Pluripotency Regulators. Plos One *6*.

Hall, R.K., Yamasaki, T., Kucera, T., Waltner-Law, M., O'Brien, R., and Granner, D.K. (2001). Regulation of phosphoenolpyruvate carboxykinase and insulin-like growth factor-binding protein-1 gene expression by insulin. The role of winged helix/forkhead proteins (vol 275, pg 30169, 2000). J Biol Chem *276*, 23212-23212.

Han, J.Y., Yuan, P., Yang, H., Zhang, J.Q., Soh, B.S., Li, P., Lim, S.L., Cao, S.Y., Tay, J.L., Orlov, Y.L., *et al*. (2010). Tbx3 improves the germ-line competency of induced pluripotent stem cells. Nature *463*, 1096-U1120.

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., *et al*. (2010). Nucleosome dynamics define transcriptional enhancers. Nat Genet *42*, 343-347.

He, S.Y., Pant, D., Schiffmacher, A., Meece, A., and Keefer, C.L. (2008). Lymphoid enhancer factor 1-mediated wnt signaling promotes the initiation of trophoblast lineage differentiation in mouse embryonic stem cells. Stem Cells *26*, 842-849.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell *38*, 576-589.

Heng, J.C., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., *et al*. (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell Stem Cell *6*, 167-174.

Hernan, R., Heuermann, K., and Brizzard, B. (2000). Multiple epitope tagging of expressed proteins for enhanced detection. Biotechniques *28*, 789-793.

Hoffmeyer, K., Raggioli, A., Rudloff, S., Anton, R., Hierholzer, A., Del Valle, I., Hein, K., Vogt, R., and Kemler, R. (2012). Wnt/beta-catenin signaling regulates telomerase in stem cells and cancer cells. Science *336*, 1549-1554.

Howard, P.K., Shaw, J., and Otsuka, A.J. (1985). Nucleotide-Sequence of the Bira-Gene Encoding the Biotin Operon Repressor and Biotin Holoenzyme Synthetase Functions of Escherichia-Coli. Gene *35*, 321-331.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols *4*, 44-57.

Huang, S.M.A., Mishina, Y.M., Liu, S.M., Cheung, A., Stegmeier, F., Michaud, G.A., Charlat, O., Wiellette, E., Zhang, Y., Wiessner, S., *et al*. (2009b). Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. Nature *461*, 614-620.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Ji, H.K., Jiang, H., Ma, W.X., Johnson, D.S., Myers, R.M., and Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat Biotechnol *26*, 1293-1300.

Jiang, J., Chan, Y.-S., Loh, Y.-H., Cai, J., Tong, G.-Q., Lim, C.-A., Robson, P., Zhong, S., and Ng, H.-H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353-360.

Kelly, K.F., Ng, D.Y., Jayakumaran, G., Wood, G.A., Koide, H., and Doble, B.W. (2011). beta-Catenin Enhances Oct-4 Activity and Reinforces Pluripotency through a TCF-Independent Mechanism. Cell Stem Cell *8*, 214-227.

Kim, J. (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. Cell *132*, 1049-1061.

Kohonen, T. (2012). Essentials of the self-organizing map. Neural Netw.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X.H., Chi, A.S., *et al*. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. Plos Genet *4*, -.

Lanner, F., and Rossant, J. (2010). The role of FGF/Erk signaling in pluripotent cells. Development *137*, 3351-3360.

Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.L., *et al*. (2008). Germline competent embryonic stem cells derived from rat blastocysts. Cell *135*, 1299-1310.

Lindsley, R.C., Gill, J.G., Kyba, M., Murphy, T.L., and Murphy, K.M. (2006). Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. Development *133*, 3787-3796.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y., *et al*. (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*.

Lluis, F., Pedone, E., Pepe, S., and Cosma, M.P. (2008). Periodic Activation of Wnt/beta-Catenin Signaling Enhances Somatic Cell Reprogramming Mediated by Cell Fusion. Cell Stem Cell *3*, 493-507.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al*. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

Lyashenko, N., Winter, M., Migliorini, D., Biechele, T., Moon, R.T., and Hartmann, C. (2011). Differential requirement for the dual functions of beta-catenin in embryonic stem cell self-renewal and germ layer formation. Nat Cell Biol *13*, 753-761.

Marson, A., Foreman, R., Chevalier, B., Bilodeau, S., Kahn, M., Young, R.A., and Jaenisch, R. (2008a). Wnt signaling promotes reprogramming of somatic cells to pluripotency. Cell Stem Cell *3*, 132-135.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., *et al*. (2008b). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Mason, M.J., Plath, K., and Zhou, Q. (2010). Identification of context-dependent motifs by contrasting ChIP binding data. Bioinformatics *26*, 2826-2832.

McKay, M.M., and Morrison, D.K. (2007). Integrating signals from RTKs to ERK/MAPK. Oncogene *26*, 3113-3121.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al*. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Merrill, B.J., Pasolli, H.A., Polak, L., Rendl, M., Garcia-Garcia, M.J., Anderson, K.V., and Fuchs, E. (2004). Tcf3: a transcriptional regulator of axis induction in the early embryo. Development *131*, 263-274.

Mikkelsen, T.S., Ku, M.C., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al*. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-U552.

Morkel, M., Huelsken, J., Wakamiya, M., Ding, J.X., van de Wetering, M., Clevers, H., Taketo, M.M., Behringer, R.R., Shen, M.M., and Birchmeier, W. (2003). beta-Catenin regulates Cripto- and Wnt3-dependent gene expression programs in mouse axis and mesoderm formation. Development *130*, 6283-6294.

Murray, J.T., Campbell, D.G., Morrice, N., Auld, G.C., Shpiro, N., Marquez, R., Peggie, M., Bain, J., Bloomberg, G.B., Grahammer, F., *et al*. (2004). Exploitation of KESTREL to identify NDRG family members as physiological substrates for SGK1 and GSK3. Biochem J *384*, 477-488.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Gene Dev *12*, 2048-2060.

Nusse, R. (2005). Wnt signaling in disease and in development. Cell Res *15*, 28-32.

Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D., and Niwa, H. (2006). Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. Biochem Bioph Res Co *343*, 159-166.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Ring, D.B., Johnson, K.W., Henriksen, E.J., Nuss, J.M., Goff, D., Kinnick, T.R., Ma, S.T., Reeder, J.W., Samuels, I., Slabiak, T., *et al*. (2003). Selective glycogen synthase kinase 3 inhibitors potentiate insulin activation of glucose transport and utilization in vitro and in vivo. Diabetes *52*, 588-595.

Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S., and Polakis, P. (1996). Binding of GSK3 beta to the APC-beta-catenin complex and regulation of complex assembly. Science *272*, 1023-1026.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med *10*, 55-63.

Schaniel, C., Ang, Y.-S., Ratnakumar, K., Cormier, C., James, T., Bernstein, E., Lemischka, I.R., and Paddison, P.J. (2009). Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. Stem Cells *27*, 2979-2991.

Schatz, P.J. (1993). Use of Peptide Libraries to Map the Substrate-Specificity of a Peptide-Modifying Enzyme - a 13 Residue Consensus Peptide Specifies Biotinylation in Escherichia-Coli. Bio-Technol *11*, 1138-1143.

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., and Orkin, S.H. (2008). EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. Mol Cell *32*, 491-502.

Singla, D.K., Schneider, D.J., LeWinter, M.M., and Sobel, B.E. (2006). wnt3a but not wnt11 supports self-renewal of embryonic stem cells. Biochem Bioph Res Co *345*, 789-795.

Skarnes, W.C., Kelly, O.G., and Pinson, K.I. (2004). The Wnt co-receptors Lrp5 and Lrp6 are essential for gastrulation in mice. Development *131*, 2803-2815.

Taelman, V.F., Dobrowolski, R., Plouhinec, J.L., Fuentealba, L.C., Vorwald, P.P., Gumper, I., Sabatini, D.D., and De Robertis, E.M. (2010). Wnt signaling requires sequestration of glycogen synthase kinase 3 inside multivesicular endosomes. Cell *143*, 1136-1148.

Takada, R., Satomi, Y., Kurata, T., Ueno, N., Norioka, S., Kondoh, H., Takao, T., and Takada, S. (2006). Monounsaturated fatty acid modification of Wnt protein: Its role in Wnt secretion. Developmental Cell *11*, 791-801.

Takada, S., Mcmahon, J., Stark, K., and Mcmahon, A. (1994). Wnt-Genes in the Regulation of Mesodermal Development. J Cell Biochem, 465-465.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663-676.

Tang, Q., Chen, Y., Meyer, C., Geistlinger, T., Lupien, M., Wang, Q., Liu, T., Zhang, Y., Brown, M., and Liu, X.S. (2011). A Comprehensive View of Nuclear Receptor Cancer Cistromes. Cancer Res.

Ten Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., Siu, R.K., and Nusse, R. (2011). Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. Nat Cell Biol *13*, 1070-U1088.

Tenzen, T., Allen, B.L., Cole, F., Kang, J.-S., Krauss, R.S., and McMahon, A.P. (2006). The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. Developmental Cell *10*, 647-656.

Tucker, K.L., Wang, Y.K., Dausman, J., and Jaenisch, R. (1997). A transgenic mouse strain expressing four drug-selectable marker genes. Nucleic Acids Res *25*, 3745-3746.

Vokes, S.A., Ji, H., Mccuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh, W.J.R., Davidson, E.H., Wong, W.H., and Mcmahon, A.P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development *134*, 1977-1989.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684-687.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol *13*, 838-845.

Wu, C.I., Hoffman, J.A., Shy, B.R., Ford, E.M., Fuchs, E., Nguyen, H., and Merrill, B.J. (2012). Function of Wnt/beta-catenin in counteracting Tcf3 repression through the Tcf3-beta-catenin interaction. Development *139*, 2118-2129.

Yamaguchi, T.P., Takada, S., Yoshikawa, Y., Wu, N.Y., and McMahon, A.P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. Gene Dev *13*, 3185-3190.

Yi, F., Pereira, L., Hoffman, J.A., Shy, B.R., Yuen, C.M., Liu, D.R., and Merrill, B.J. (2011). Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nat Cell Biol *13*, 762-770.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008a). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res *21*, 1273-1283.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W*., et al*. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol *9*, -.

Chapter 3

Genome-wide study of Shh-directed neural patterning in mouse ventral neural tube

Addendum

A portion of this chapter is published as:

Peterson KA, Nishi Y, Ma W, Vedenko A, Shokri L, Zhang X, McFarlane M, Baizabal JM, Junker JP, van Oudenaarden A, Mikkelsen T, Bernstein BE, Bailey TL, Bulyk ML, Wong WH, McMahon AP Neural-specific Sox2 input and differential Gli-binding affinity provide context and positional information in Shh-directed neural patterning. Genes Dev. 2012 Dec 15;26(24):2802-16. doi: 10.1101/gad.207142.112.

YN, KAP and APM conceived the study and designed the experiments. YN and KAP collaborated on the experimental part of the Gli1/Sox2 project. WM and KAP established the genomic data analysis pipeline, and the initial processing of genomic data. XZ performed the data integration of Gli1 and Sox2 ChIP-seq with microarray expression data. YN performed the ChIP-seq against Olig2, Nkx2.2, Nkx6.1, and temporal histone modifications as well as RNA-seq in the *in vitro* differentiation system of generating neural progenitor cells from mESCs. XZ analyzed the ChIP-seq and RNA-seq data, and performed the analysis associated with Olig2 and DNA recognition specificity in pMN cell differentiation.

Abstract

In the developing vertebrate neural tube, Sonic Hedgehog (Shh) signaling acts in a time- and concentration-dependent manner to activate position-specific ventral programs of neural progenitor specification. Gli proteins are transcriptional regulators that act at the interface between Shh signaling and its gene targets to specify six different ventral cell populations. Subsequently, these populations are further refined and sharpened via cross-repression by ventral cell determinants. This study seeks to understand how dynamic Shh signaling governs a complex patterning response by focusing on the Gli factors, the pan-neural factor Sox2 and their transcriptional targets. We highlight Sox2 as an early delineator of the neural lineage by switching on Shh-responsive ventral neural patterning together with Gli1. Furthermore, by assessing *in silico* DNA binding specificity of different Olig2 dimer species, we also propose a model that describes how Olig2 mediates specification of oligodendrocytes progenitor cell fate from motor neuron progenitors mediated by Olig2.

Introduction

Pan-neural factor Sox2 and neural priming

The secreted signaling molecule Sonic hedgehog (Shh), a member of the Hh family, is a long-range morphogen that patterns many embryonic structures including the VNT, limb bud and somites (Ingham and McMahon, 2001)(Figure 1.1C). How this broadly utilized signaling pathway mediates tissue-specific regulatory responses is a general question of interest.

Recent studies have started to address specific outcomes through broadly utilized signaling processes. A cell-type specific response may be established by promoting chromatin-remodeling activities that enable cell-type specific access to a transcriptional component regulated by a given pathway. The Zon and Young groups' work （2011） has demonstrated that a master TF defines lineage-specific target genes, and the action of this factor is followed by binding of transcriptional effectors of specific signaling pathways (Mullen et al., 2011; Trompouki et al., 2011). Similarly, Fox factors are known to modify local epigenetic status to generate TF accessible sites. Cell-type specific response can also be established via a passive mechanism, in which a TF can prevent nucleosome repositioning and serve as a placeholder to facilitate the binding of a second TF. In both mechanisms, the action of the first regulatory factor determines the subset of target sites in the genome that can be bound by the second TF. The action of both factors is required to activate a given regulatory program in target cells.

Shh morphogen signaling establishes an activity gradient of the Gli family of TFs along the dorsal-ventral (D-V) axis of the VNT. Gli factor binding to DNA dictates the transcriptional response to Shh signaling. As elaborated above, the neural-specific response to Shh signaling could result from differential accessibility of Gli factors to cell-type specific enhancers. However, previous work from the McMahon lab on the Gli3 directed GRN in mammalian limb patterning

showed that almost all verified neural, *cis*-regulatory elements are bound by Gli3 which acts largely to silence Hedgehog target genes. However, when Gli3 is removed neural targets are not inappropriately activated within the limb. These results argue for context specific inputs acting in conjunction with Gli factors. In the context of the developing CNS, general neural identity genes may act in concert with Gli factors to ensure an appropriate cell type specific response to Shh.

Of the three classes of transcriptional activators that are broadly expressed in early neuronal cells, Pax, Pou, and Sox, only Sox family factors are expressed prior to Shh signaling activation (Peterson et al., 2012; Scotting and Rex, 1996). Furthermore, expression of a Sox-engrailed repressor fusion leads to a loss of neural identity consistent with Sox genes playing a critical early role in neural identity (Kishi et al., 2000). Within the Sox family, SoxB1 proteins (Sox1, Sox2, and Sox3) are expressed in most neural progenitors. Sox2 is present within neural cells prior to Shh signaling, while Sox1 is activated in conjunction with Shh signaling, and Sox3 is expressed in neural precursor cells (Bergsland et al., 2011). Sox2 level has been shown to modulate neural ectodermal differentiation over mes-endoderm differentiation with opposing roles to Oct4 during ESC early differentiation (Thomson et al., 2011). Sox2 is therefore the most likely pioneer Sox factor to mediate the initial neural specific response to Shh signaling in the NT.

bHLH Protein Olig2 and differentiation of pMN into motor neuron and oligodendrocytes progenitors

In the VNT, most cell type determinants are repressors, raising the question of how Olig2 repressor activity specifies a MN progenitor state that when Olig2 activity is removed, cells then differentiate into MN or OLP. In addition, by over-expressing Pax6, Olig2, or Nkx2.2 and

examining the faction of differentiated cells within the population, Sugimori et al. (2007) showed that these patterning factors can bias certain fates, but in general cannot promote differentiation (Sugimori et al., 2007). To examine these issues, I have focused on Olig2--a key regulator of motor neuron (MN) and oligodendrocytes progenitor (OLP) fate specification which arise at different times from a shared progenitor pool.

The bHLH TF Olig2 was first discovered to have a deterministic role in OLP specification (Zhou et al., 2001; Zhou et al., 2000), but later was found to have an earlier role in MN precursor generation (Lu et al., 2002; Nieto et al., 2001; Zhou and Anderson, 2002). Olig2 expression is restricted to MN progenitors in the ventricular zone of the early VNT. Later in development, Olig2-expressing cells encompassing the early MN progenitor domain switch to the production of OLP (Sugimori et al., 2007; Zhou and Anderson, 2002). Thus, the pMN Olig2 population undergoes sequential programs of neurogenesis and gliogenesis, which generate all of the MNs and ~80% of the OLPs in mouse spinal cord (Li et al., 2011). The pMN domain is bounded ventrally by an Nkx2.2-expressing pV3 interneuron progenitor domain and dorsally by an Irx3-expressing pV2 interneuron progenitor domain. Repression by Nkx2.2 and Irx3 is critical for delimiting the Olig2 domain. In mouse embryos lacking both Olig1 and Olig2, two structurally and functionally similarly bHLH family proteins, both MN and OLP differentiation are largely eliminated (Lu et al., 2002; Zhou and Anderson, 2002). The animals also exhibited ventral expansion of the pV2 domain and a switch of cell fate from MN to pV2 interneurons and astrocytes.

Where it has been best studied in MN specification, Olig2 functions as a transcriptional repressor (Mizuguchi et al., 2001; Novitch et al., 2001). An EnR-Olig2 fusion construct in which the engrailed repressor domain was fused to the DNA binding bHLH domain of Olig2 mimicked

the patterning activities of the full-length Olig2, and induced ectopic MN generation on electroporation in the chick spinal cord. Conversely, expression of an Olig2 fusion construct with the transcriptional activation domain of the viral protein herpes simplex virus protein 16 (Olig2-VP16) reduced MN generation (Mizuguchi et al., 2001; Novitch et al., 2001).

Olig2 functions alongside other bHLH factors to determine the initiation of neurogenesis, neuronal subtype specification, and the MN-to-OLP switch (Li et al., 2011). These factors are divided into two major groups: proneural TFs that promote terminal differentiation, and anti-neuronal TFs that promote progenitor fate maintenance (Castro et al., 2006; Nieto et al., 2001; Zhou et al., 2000) based on there cellular activities; the timing of neurogenesis is determined by proneural and anti-neuronal bHLH proteins. bHLH proteins can also be classified as repressor or activator type according to their transcriptional activity. Hes family factors contribute to the repressor-type represented and Ascl1, Math, and Neurogenin to the activator-type.

For example, Neurog2 (Neurogenin2) and Ascl1 (achaete-scute complex homolog-1) are two major proneural bHLH proteins expressed in progenitor cells in the ventral ventricular zone with distinct temporal features. Expression of Neurog2 correlates with neurogenesis, during which it partly overlaps with Olig2 expression and is turned off before oligodendrogenesis. In contrast, Ascl1 is co-expressed with Olig2 in pMN from the beginning of OLP generation, after Neurog2 expression is lost. The fact that Neurog2:Ascl1 double-mutant mice exhibit a reduction of neurogenesis and a premature generation of astrocytic precursors with a comprise in OLPs suggests a critical role of Neurog2 and Ascl1 in neurogenesis and oligodendrogenesis (Nieto et al., 2001). In contrast, in mice deficient for Hes1 or Hes5, an anti-neuronal bHLH family under Notch pathway, premature neural differentiation occurs (Kageyama and Nakanishi, 1997; Kageyama and Ohtsuka, 1999; Kageyama et al., 2008). Other studies have suggested synergistic

or antagonistic roles of proneural proteins and Olig2 in neuronal subtype specification using chick electroporation or *in vitro* biochemical assay (Lee et al., 2005). However, these studies are mostly restricted to analysis of individual genes, and are inadequate to formulate a general mechanism for the regulation by proneural factors, and Olig2 in cell fate specification programs.

bHLH proteins also have unique DNA sequence specificity. bHLH proteins bind to chromatin as either hetero- or homodimers. The consensus sequence is CANNTG (Murre et al., 1989). Within the bHLH factors, different family has slightly different preference to the center two nucleotides or a modified E-box six-mer. For example, Hes factors have a higher affinity to CACNAG consensus than to E-box consensus. Altering dimerization partners has been proposed as a mechanism underlying tissue-specific cell fate determination by Olig2 (Li et al., 2011). Neurog2 has been shown to form a heterodimer with Olig2 during MN differentiation (Mizuguchi et al., 2001; Novitch et al., 2001). Li et al. (2011) proposed that a phosphorylated Olig2 monomer could form homodimers and repress OLP gene expression, while heterodimers with E47 activate MN specific gene expression. Later in development, Olig2 is dephosphorylated and interacts with different co-activator bHLH factors to direct OLP specific gene expression. E47 has been shown to interact with Olig2 to directly activate the Sox10-U2 promoter in the oligodendrocytes lineage (Kuspert et al., 2011). Interestingly, Nkx2.2 has also been proposed as a potential dimerization partner of Olig2 during OLP lineage specification (Sun et al., 2003), when their expression domains overlap during OLP differentiation (Zhou et al., 2001).

While these studies have provided a solid foundation for understanding the role of Olig2 in MN-OLP fate determination, their focus on extrapolating insights from single genomic loci begs the question of the generality of these regulatory modalities.

In this study, we showed that Sox2 is a context-specific determinant of the Shh response in the NT. We found that assessing genomic footprints of both Sox2 and Gli1 allows for the identification of Shh responsive ventral determinants in the neural context. We proposed the role of Sox2 in transcriptional regulation of Shh target genes by promoting chromatin remodeling that enhances promoter accessibility. In addition, to understand the regulatory mechanisms employed by repressors, we computationally explored the binding specificity of Nkx2.2, Olig2, and Nkx6.1. We found that these factors interact directly with the DNA and share a common set of *cis*-elements in many cases. Furthermore, their expression patterns overlap extensively with that of Sox2, suggesting the possibility for generating active input for the repressor *cis*-elements. Lastly, we propose that in pMN cell differentiation, Olig2 directs different cell fates by forming various dimer species with distinct chromatin-binding properties.

Results

Gli1 and Sox2 chromatin co-binding in neural progenitors denote actively transcribed genes with Shh responsiveness

Specification of Shh-dependent neural progenitor populations within the VNT occurs in a time- and dosage-dependent manner (Fig. 1.1C). To further elucidate the transcriptional regulatory program downstream from Shh signaling in neural progenitors, we used an *in vitro* model of Shh-dependent patterning (Fig. 1.1D)(Wichterle et al., 2002).

Shh pathway activation was achieved by exposing *in vitro*-derived neural progenitors to Smo agonist (SAG), a small molecule that acts as a potent activator of the pathway (Chen et al., 2002). Immunofluorescence characterization of SAG-treated neural progenitors highlighted both the time- and concentration-dependent responses of Class II target genes (Figure 3.1A). Probably

one of the biggest concerns regarding the usage of *in vitro* differentiation system to study in vivo neuronal development is the similarity between the two. We compared the temporal progression of Shh-dependent cell-types *in vivo* to those observed in an *in vitro* model of Shh patterning. SAG-dependent induction of Class II Shh targets closely followed the *in vivo* time-course of neutral patterning: Foxa2 was detected at 24h, Olig2 and Nkx6.1 at 36h, and Nkx2.2 at 48h (Figure 3.1B). Overall, we see a similar feature in the pattern of target gene activation via immunofluorescence, although we do notice that the *in vitro* response is much slower in general. In addition, several of the identified *cis*-regulatory elements from the *in vitro* system have been verified by G0 transgenic assay for enhancer elements. Our analysis has focused on a one time-point after SAG treatment, which is chosen to maximize the diversity of Shh-dependent cell types.

Genome-wide location analysis of Gli1 DNA association in ESC-derived neural progenitors was performed using a Gli1Flag transgene (Vokes et al., 2007). We defined a prioritized set of 841 enriched Gli1-binding regions (GBRs) by intersecting ChIP-seq data independently verified in biological replicates (false discovery rate [FDR] = 0.01; CisGenome peak calling algorithm) (Figure 3.1C-G: Gli1 peaks in genome browser) (Ji et al., 2008). A small number of developmental signals including Shh are utilized in multiple tissue types but their action in each tissue triggers tissue context-dependent GRNs. So this raised the question that how neural-specific Shh responses are initiated without activating Shh responses in other tissues? We hypothesize that transcriptional activators that are expressed prior to Shh signaling in neuronal cells might serve to prime a neural-specific responses. Recent studies have shown that SoxB1 family TF Sox2 is expressed within neural cells prior to Shh signaling (Bergsland et al., 2011), while other neuron expressed transcription activators are expressed after Shh signaling

Figure 3.1

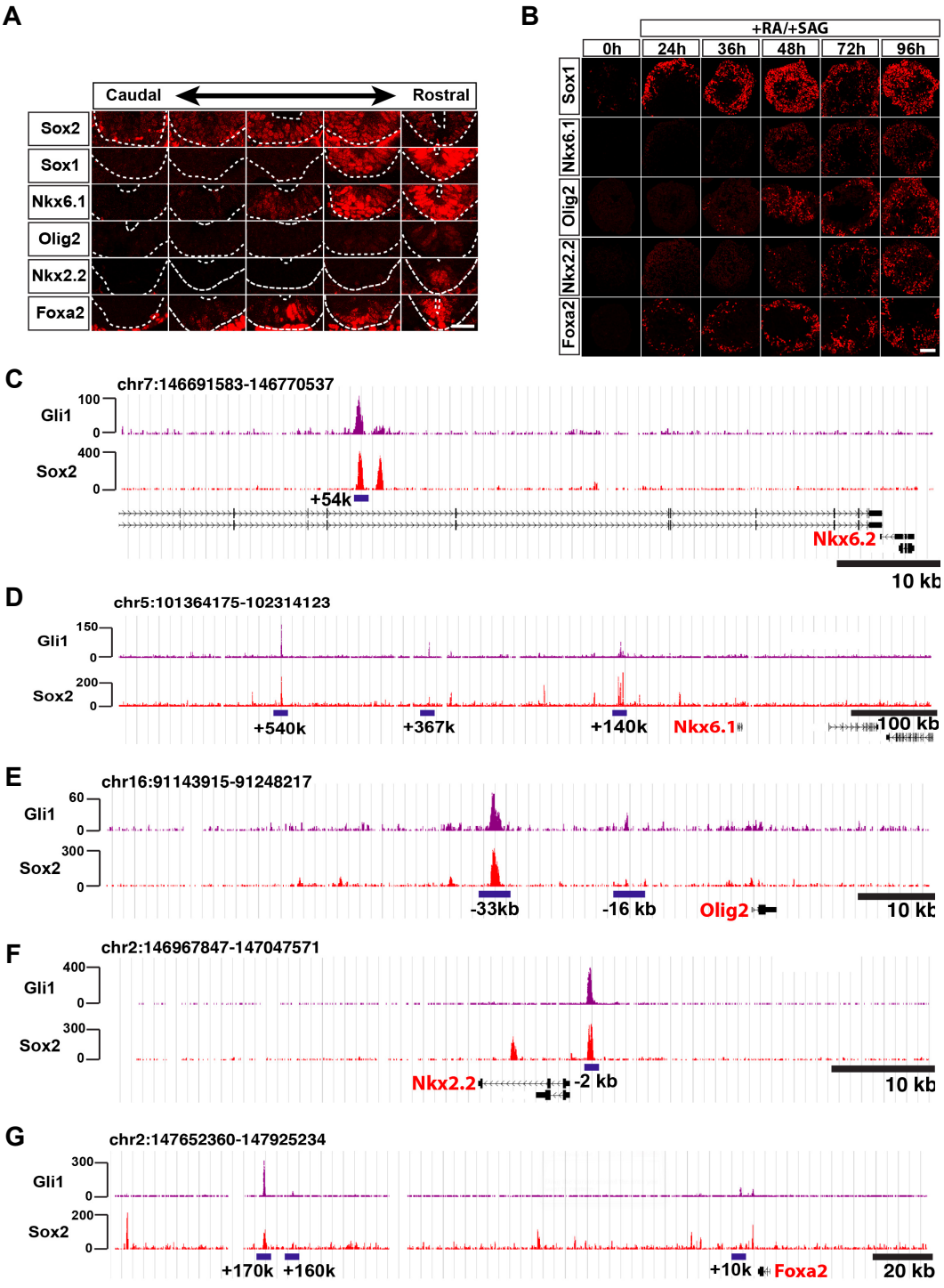
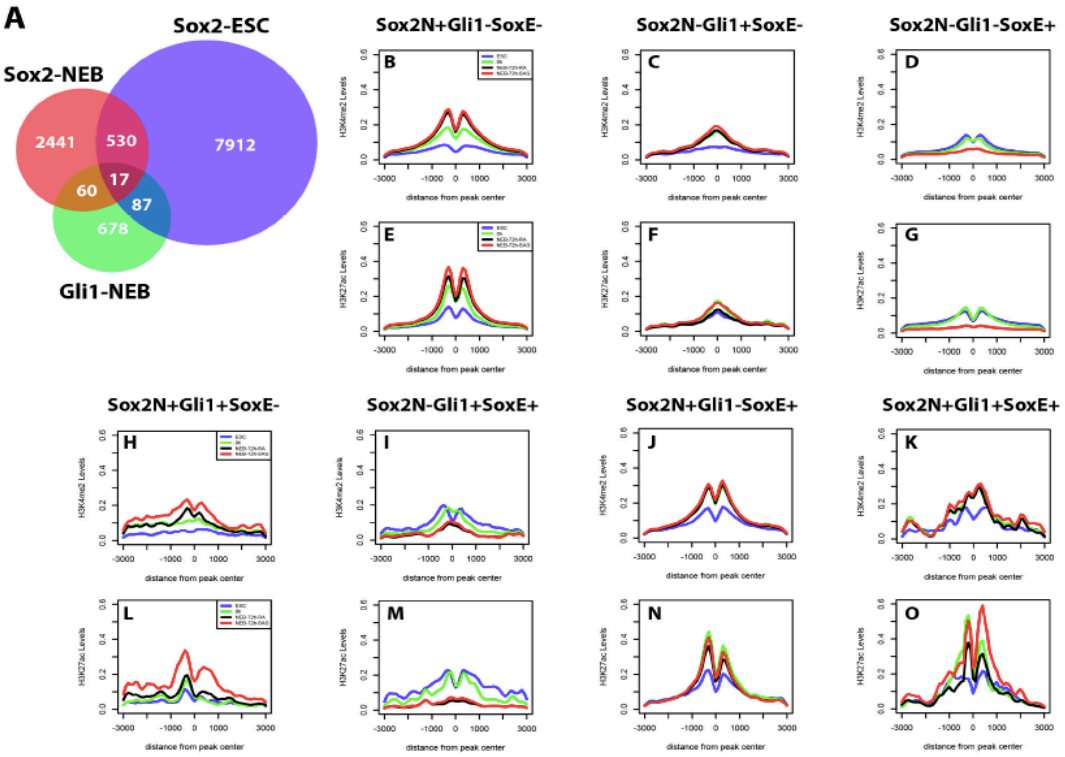

A
Caudal
Rostral
Sox2
Sox1
Nkx6.1
Olig2
Nkx2.2
Foxa2
B
+RA/+SAG
0h
24h
36h
48h
72h
96h
Sox1
Nkx6.1
Olig2
Nkx2.2
Foxa2
C
chr7:146691583-146770537
Gli1
100
0
Sox2
400
0
+54k
Nkx6.2
10 kb
D
chr5:101364175-102314123
Gli1
150
0
Sox2
200
0
+540k
+367k
+140k
Nkx6.1
100 kb
E
chr16:91143915-91248217
Gli1
60
0
Sox2
300
0
-33kb
-16 kb
Olig2
10 kb
F
chr2:146967847-147047571
Gli1
400
0
Sox2
300
0
Nkx2.2
-2 kb
10 kb
G
chr2:147652360-147925234
Gli1
300
0
Sox2
200
0
+170k
+160k
+10k
Foxa2
20 kb

Figure 3.1 (Continued)
The use of in vitro differentiation system for neural patterning studies.
(A) Immunofluorescence characterization of key neural markers in E8.5 neural tube. Scale bar: 20μm.
(B) Time-dependent emergence of ventral cell populations in neural embryoid bodies (NEBs) treated with retinoic acid (RA) and Hh pathway agonist (SAG). Scale bar: 50μm.
(C) –(G) CisGenome browser demonstration of Sox2 and Gli1 common regulatory elements in neural progenitors. Gli1 (replicate A, purple) and Sox2 (red) ChIP-seq signal is shown. Blue underline denotes Gli1 binding signal that passed peak detection threshold with relative distance to TSS shown below.

activation. Sox2 is therefore the most likely pioneer Sox factor to mediate the initial neural specific response to Shh signaling in the NT. To examine the potential regulatory role for SoxB1 family input to Gli1-bound CRMs, we performed Sox2 ChIP-seq analysis in neural progenitors. All the data on chromatin were collected from ES cell generated neural responses. Thus, features of the chromatin status may reflect *in vitro* generated or ES cell-derived signatures. To exclude noise from ESC signature, I integrated Sox2 ChIP-seq from mESC (Marson et al., 2008). I intersected ChIP-seq data of Gli1 in neural derivatives with that of Sox2 in neural derivatives and excluded Sox2 peaks in mESCs, and ~12% of GBRs overlap specifically with neural-associated Sox2-bound regions (p-value <1.3e-50, Fisher Exact test), with only a small number of these also bound by Sox2 in ESCs (Figure 3.2A).

To assess how Sox2 binding influences Gli1-bound CRM activity, we examined the chromatin state of GBRs using H3K4me2 as a general mark for enhancers and H3K27ac for active enhancers (Barski et al., 2007; Creyghton et al., 2010; Roh et al., 2006). ChIP-seq profiling of H3K4me2 and H3K27ac at GBRs associated with Sox2N (Sox2 binding in NEB, abbrev: Sox2N) showed higher H3K4me2 levels than those observed in their non-Sox2N bound counterparts suggesting that Sox2 binding enhances activity of Gli-mediated CRMs (Figure 3.2B-O). Specifically, regions bound by Gli1 but not Sox2N (Gli1+ Sox2N-) showed moderate levels of H3K4me2 and H3K27ac independent of Shh pathway stimulation (Figure 3.2C, F, I, M). Regions bound by Sox2N but not Gli1 (Gli1- Sox2N+) showed significant levels for both H3K4me2 and H3K27ac marks in a largely Shh-independent manner (Figure 3.2B, E, J, N). On the other hand, regions bound by both Gli1 and Sox2N (Gli1+ Sox2N+) showed overall high levels of H3K4me2 and H3K27ac signals that were further enhanced upon Shh pathway activation (Figure 3.2H, L, K, O). Meta-site analysis of H3K4me2 and H3K27ac profiles

Figure 3.2


A
Sox2-ESC
Sox2-NEB
Gli1-NEB
2441
530
7912
60
17
87
678
Sox2N+Gli1-SoxE-
Sox2N-Gli1+SoxE-
Sox2N-Gli1-SoxE+
Sox2N+Gli1-SoxE-
Sox2N-Gli1+SoxE+
Sox2N+Gli1-SoxE+
Sox2N+Gli1+SoxE+
B
C
D
E
F
G
H
I
J
K
L
M
N
O
distance from peak center

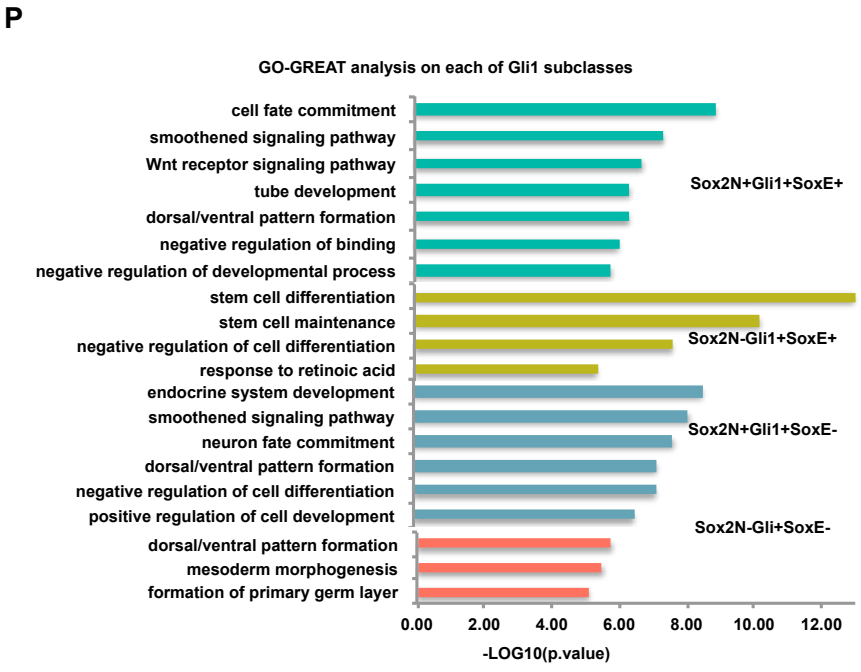

P
GO-GREAT analysis on each of Gli1 subclasses
cell fate commitment
smoothened signaling pathway
Wnt receptor signaling pathway
tube development
dorsal/ventral pattern formation
negative regulation of binding
negative regulation of developmental process
Sox2N+Gli1+SoxE+
stem cell differentiation
stem cell maintenance
negative regulation of cell differentiation
response to retinoic acid
Sox2N-Gli1+SoxE+
endocrine system development
smoothened signaling pathway
neuron fate commitment
dorsal/ventral pattern formation
negative regulation of cell differentiation
positive regulation of cell development
Sox2N+Gli1+SoxE-
dorsal/ventral pattern formation
mesoderm morphogenesis
formation of primary germ layer
Sox2N-Gli+SoxE-
0.00
2.00
4.00
6.00
8.00
10.00
12.00
-LOG10(p.value)

Figure 3.2 Continued

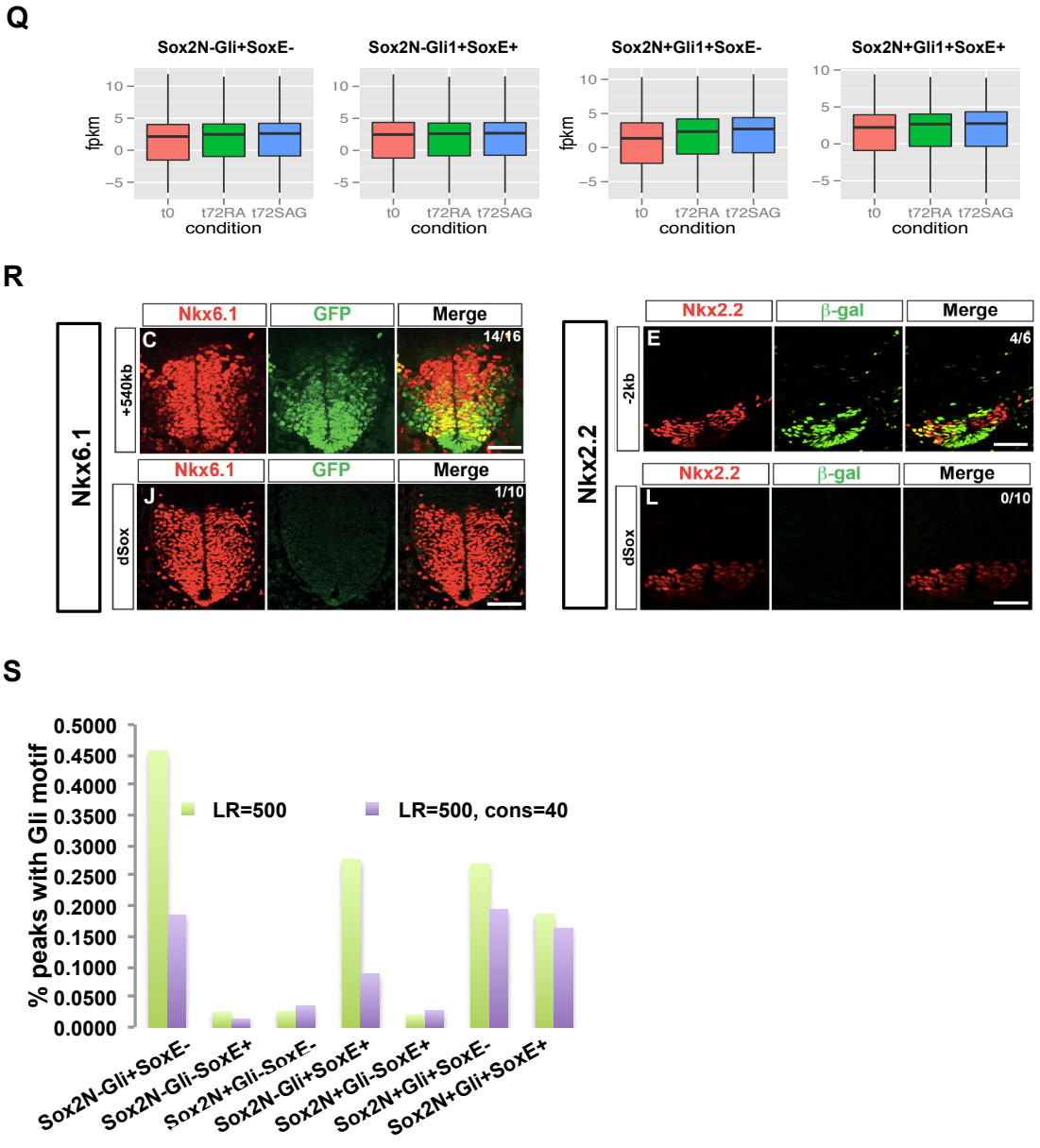


Figure 3.2 (Continued)
Sox2 binding might prime cis-regulatory elements by opening up chromatin structure prior to repressor binding.
(A) Venn diagram of Sox2 ESC, Sox2 NEB, and Gli1 NEB binding data.
(B)-(O) Summary of histone modifications (H3K4me2 and H3K27ac) corresponding different portions of the Venn diagram. In our in vitro neural progenitor derivation protocol (Figure 3.1x), ESCs are first cultured in suspension for two days before three days of neural induction. t=0 (green) is defined as the beginning and t=72h the end of the induction. RA (red) and RA/SAG (black) induction was performed in parallel.
(P) GO Biological Process analysis of the four types of Gli1 NEB binding regions by GREAT.
(Q) Normalized FPKM values from RNA-seq are shown as box plots across the conditions.
(R) Mutation analysis of cis-elements near Nkx6.1 (+540kb) (top) and Nkx2.2 (-2kb) (bottom) indicate the necessity of Sox2 in enhancer activities. Scale bar: 50μm.
(S) Percentage of computationally predicted Gli sites at Sox2 sites that are occupied by Gli1. Two criteria are used: likelihood ratio (LR) = 500, and LR=500 plus phastcons conservation score = 40.

highlighted a characteristic dip positioned at the Sox2-binding peak center that was absent from regions bound by Gli1 alone, suggesting nucleosome displacement by Sox2 but not Gli1 (Figure 3.2B–O). In addition, from Gene Ontology (GO) analysis of peak associated genes, we found peaks co-bound by Gli1 and Sox2N are significantly more enriched for neural development categories, whereas those Gli1 bound sites that do not intersect with Sox2 in NEB are not as strongly enriched for neural development, but are enriched for other developmental processes (Figure 3.2P). This set of neural-specific co-bound regions correlated with differential gene expression. The largest increase in mean FPKM (fragments per kilo base per million reads) values from RNA-seq data were observed for genes near peaks with both Sox2 and Gli1 present (Figure 3.2Q). Comparison of GBRs to Sox2 bound regions highlighted a specific intersection with neural targets, most notably, Class II targets (*Foxa2, Nkx2.2, Nkx2.9, Olig2, Nkx6.1, and Nkx6.2*) (Figure 3.1C-G). Interestingly, this set of targets bound by Gli1 and Sox2 in neural progenitors was distinct from the set of regions bound by Sox2 in ESC. Only a limited overlap was observed between Sox2 binding regions in ESCs and neural progenitors, in agreement with a recent report (Bergsland et al., 2011). These findings highlight potential interactions between Sox2 and Gli1 where Sox2 may help to confer neural specificity to the Shh/Gli response.

Sox2 priming might be necessary for Gli1 binding in Class II gene *cis*-element prior to Shh responsiveness

Recent studies indicate that Sox2 activity is the earliest determinant of neural identity (Thomson et al., 2011), Sox2 expression precedes the Shh response *in vivo* (Figure 3.1A), and Sox activity is required to maintain a neural progenitor state (Bylund et al., 2003). In addition to temporal precedence of Sox2 expression to Shh responsiveness, Sox2 binding sites were also required for

the activities for Gli-mediated CRM (Figure 3.2R). Mutations predicted to prevent Sox2 binding in the Nkx6.1 element (+540kb) and Nkx2.2 element (-2kb) resulted in the silencing of reporter activity in 9 out of 10 and all transgenic embryos, respectively. Given previous evidence of Gli-dependent activity from Gli-directed mutagenesis (Lei et al., 2006), neither Sox2 nor Gli^A input alone is sufficient to activate the Nkx2.2 CRM. Thus these results suggest that both Sox2 and Gli1 are required for Gli-mediated Shh responsiveness.

Since only ~12% of GBRs overlap specifically with neural-associated Sox2-bound regions, Sox2 is not required for the majority of Gli1 binding. However, the question is whether Sox2 binding and Gli1 motif are sufficient for Gli1 binding, especially for GBRs near Class II determinants. If so, we expect to see Gli factor occupies a significant number of Gli motif sites in Sox2 binding peaks. We observed that close to 25% of Sox2 NEB peaks with Gli motif are bound by Gli1 (Figure 3.2S, Sox2N+Gli+SoxE- and Sox2N+Gli+SoxE+), while for Sox2 NEB peaks without Gli1 binding, the percentage of Gli sites drop to less than 5% (Figure 3.2S, Sox2N+Gli-SoxE+ and Sox2N+Gli-SoxE-). This suggests that either additional factors might be required for this set of Gli1 binding or Gli1 can bind to deviant Gli motif in the presence of Sox2, but Gli1 preferably interacts with DNA via direct binding. Together with the open chromatin histone signature at Sox2/Gli1 binding sites and the temporal precedence of Sox2 expression to Shh responsiveness, the result suggests that Sox2 might serve to enable Gli1 binding by opening up the chromatin structure, but additional factors remain to be found out that ensure Gli1 binding.

Nkx6.1, Nkx2.2 and Olig2 share a common set of target binding regions

In the VNT, Shh induces a group of transcriptional repressors (Class II target genes) to specify neural progenitor subtypes. Unraveling the regulatory actions of these repressors is an important problem. Nkx2.2 and Olig2 mark two of the most ventral progenitor populations, pV3 and pMN,

respectively, while Nkx6.1 is present in both these progenitor pools, in the ventral medial floor plate adjacent to the pV3 domain, and in more dorsal pV2 progenitor adjacent to the pMN population. (Figure 1.1B). Each is a direct target of Shh signaling, and Olig2 and Nkx2.2 are both necessary and sufficient for specification of their progenitor type.

As a first step we performed ChIP-seq on SAG-treated neuralized embryoid bodies following 72h of SAG treatment as before. Using CisGenome algorithm with FDR=0.01 and two-sample peak calling with input as controls, we identified 11747, 5551, and 4094, Olig2, Nkx2.2 and Nkx6.1 genomic binding peaks, respectively (Figure 3.3A). GO analysis of binding sites showed that each of these is highly enriched for TF genes and neural development regulators (Figure 3.3B). This is consistent with their roles in neural development and their action in progenitors at an early point in regulatory events that ultimately result in appropriate neural diversity and functional circuitry.

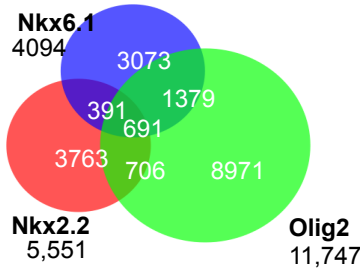
Because Nkx6.1 is co-expressed with Nkx2.2 and Olig2, and Nkx2.2 and Olig2 have mutually antagonistic properties, we intersected their binding sites to identify common and unique targets (Figure 3.3C). I saw significant overlap among the binding regions for the three factors, which indicates that a common set of *cis*-regulatory elements are employed and a core set of common target genes are co-regulated by different ventral determinants. The overlapping binding peaks might come from the same cells and be due to physically interaction between Nkx6.1 with Olig2 or Nkx6.1 with Nkx2.2. However, since Nkx2.2 and Olig2 are not co-expressed, Nkx2.2 and Olig2 binding at common *cis*-elements must come from different cells in the population, rather than co-binding. In either case, the characterization of the target genes will shed light on the functional significance of the co-regulation.

Figure 3.3

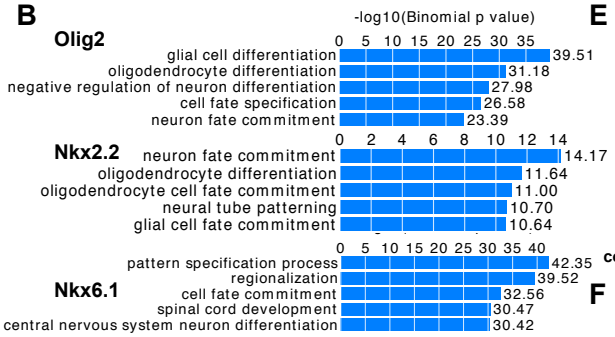
A

	ChIP		Input		
	Alignable	nrReads	Alignable	nrReads	Peaks
Nkx2-2	12,610,255	2,533,527	18,750,935	12,945,234	5,551
Nkx6-1	11,915,999	9,947,294	13,004,887	12,625,556	4,094
Olig2	10,592,204	7,018,928	14,464,709	12,538,261	11,747

C



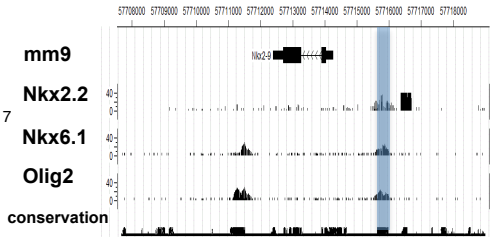
B



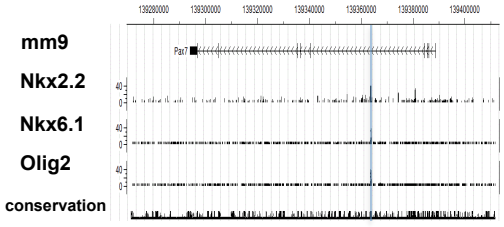
D

Genes	Peak Distance to TSS
progenitor determinants	
Nkx2-9	(-1,594)
Foxa2	(+170,290), (+216,220)
Nkx6-1	(+74,028), (+126,777), (+140,503), (+145,303), (+167,585), (+232,222)
Irx3	(-328,720), (-65,779), (+275,854), (+292,111), (+435,327)
Pax3	(+55,251)
Pax6	(-64,092)
Pax7	(+25,285)
Hh pathway genes	
Gli2	(+6,030)
Boc	(+27,762)
Gas1	(+6,397)
Ptch1	(-154,309), (+10,780)
Hhip	(+3,612)
neural subtype determinants	
Evx1	(+8,193), (+55,455), (+254,038)
Isl1	(+613,245)
Tlx3	(-30,619)
Foxd3	(-152,991), (-82,034)
Other neural genes	
Phox2b	(-3,377), (-1,884)
Vsx1	(-47,054)
Vsx2	(-81,983)

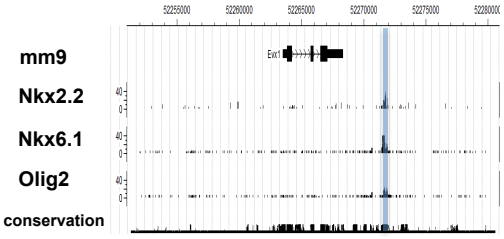
E Nkx2-9 -1594bp chr12:57707660-57718994



F Pax7 +25285bp chr4:139271152-139411726



G Evx1 +8139bp chr6:52251284-52280578



H Hhip +3612bp chr8:82535828-82627986

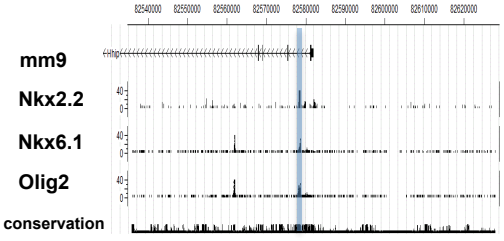


Figure 3.3 (Continued)
Nkx2.2, Olig2, and Nkx6.1 ChIP-seq reveals ventral progenitor cell maintenance by preventing alternative progenitor cell fates and post-mitotic cell differentiation.
(A) Summary statistics of ChIP-seq from and peak calling by CisGenome using FDR<0.01.
(B) GO analysis of Olig2, Nkx2.2, and Nkx6.1 peaks and top enriched terms are presented.
(C) 3-way Venn diagram shows significant overlap of binding regions among the three factors.
(D) Representative genes regulated by Nkx2.2/Olig2/Nkx6.1 from GREAT.
(E) –(H) CisGenome Browser example of the four types of target genes showing common cis-element of Nkx2.2, Olig2, and Nkx6.1. Common cis-elements are highlighted by light blue squares.

Therefore I did a GO analysis of the Nkx2.2+Olig2+Nkx6.1+ peaks using GREAT (McLean et al., 2010), and pooled out representative genes within the top enriched GO terms (Figure 3.3D). Firstly, repressors cross-repress Class-II ventral cell determinants, such as Foxa2, Nkx2.9, Nkx6.1 (Figure 3.3E). TF Binding near determinants not expressed in the same domain might suggest the cross-repression mechanism to prevent alternative cell fate, such as the repression of pMN fate by Nkx2.2 binding in pV3 progenitor cells. However, for binding near genes expressed in the same domain, such as Nkx2.2 binding near Nxk2.9 and Nkx6.1, the specific outcome remains to be determined, and one potential reason is to fine-tune expression level of the other. Secondly, repressors binding near Class-I dorsal progenitor cell markers, such as Irx3, Pax3, Pax7, and Pax6, might serve to inhibit dorsal program (Figure 3.3F). Thirdly, repressors also co-regulate neuronal subtype cell determinants, such as Isl1, Evx1, Tlx3, Foxd3, Phox2b, and Vsx1/2 (Figure 3.3G). This might contribute to the maintenance of progenitor cell state by preventing post-mitotic differentiation. Alternatively, this could represent direct exclusion of other cell fate programs by repressing all other possible neural identity determinants. Fourthly, repressors also bind to Hh signaling pathway components, such as Hhip, Ptch1, Boc, Gas1, and Gli2. For example, Hhip encodes a highly conserved, vertebrate-specific inhibitor of Hh signaling. The observation that ventral repressors interact with a *cis*-element 3.6kb upstream of Hhip promoter suggests a potential feedback loop and fine-tuning of Hh responses (Figure 3.3H). Therefore, Nkx2.2, Olig2 and Nkx6.1 are engaged in a broad spectrum of repressor activities by direct chromatin interaction with a common set of *cis*-regulatory elements, possibly by preventing alternative progenitor cell fates and post-mitotic differentiation to maintain their corresponding progenitor cell types.

Active enhancer signature characterizing Olig2 NEB binding regions

Because ventral determinants Nkx2.2, Olig2 and Nkx6.1 are all repressors, we are surprised to identify H3K4me2 +H3K27ac+ active chromatin signatures at the *cis*-elements. For example, in comparison with the matched control regions, we observed H3K4me2 and H3K27ac enrichment near Olig2 peak center (Figure 3.4A-B). The dip of H3K27ac signature in the center of Olig2 binding sites represents the displacement of histone modifications by TF binding. One caution in interpreting the data is that NEB is a mixed population of cells from all ventral cell progenitors, and thus the histone modification profiles will be an averaged signature, while in comparison, Olig2 peaks will be profiles only from pMN cells. Note that a significant contribution of the active signature might come from Sox2 from the extensive intersection of Sox2 with Olig2 in neural progenitors (Figure 3.4C). In summary, this result suggests that there is significant active input at Olig2 binding regions, and it is likely coming from activators other than Sox2.

The first candidate we tested was Sox2, which utilizes lots of common *cis*-elements with Olig2 (Figure 3.4C). An even stronger enrichment of active histone signatures were observed when we compare Olig2+Sox2+ peaks with Olig2+Sox2- peaks, although Olig2+Sox2- peaks also show a low level of H3K4me2 and H3K27ac (Figure 3.4D). Thus Sox2 activator input is likely to account for at least partially the active histone signatures in Olig2 binding regions. However, it is unlikely to be the source for maintaining mitotically active progenitor, because Sox2 expression is present within neural cells prior to Shh signaling and is gradually turned down with differentiation. Indeed, from GO analysis of Olig2+Sox2+ versus Olig2+Sox2- peaks (Figure 3.4E-F), we noticed that Olig2+Sox2+ regions seem to regulate genes expressed at an

Figure 3.4

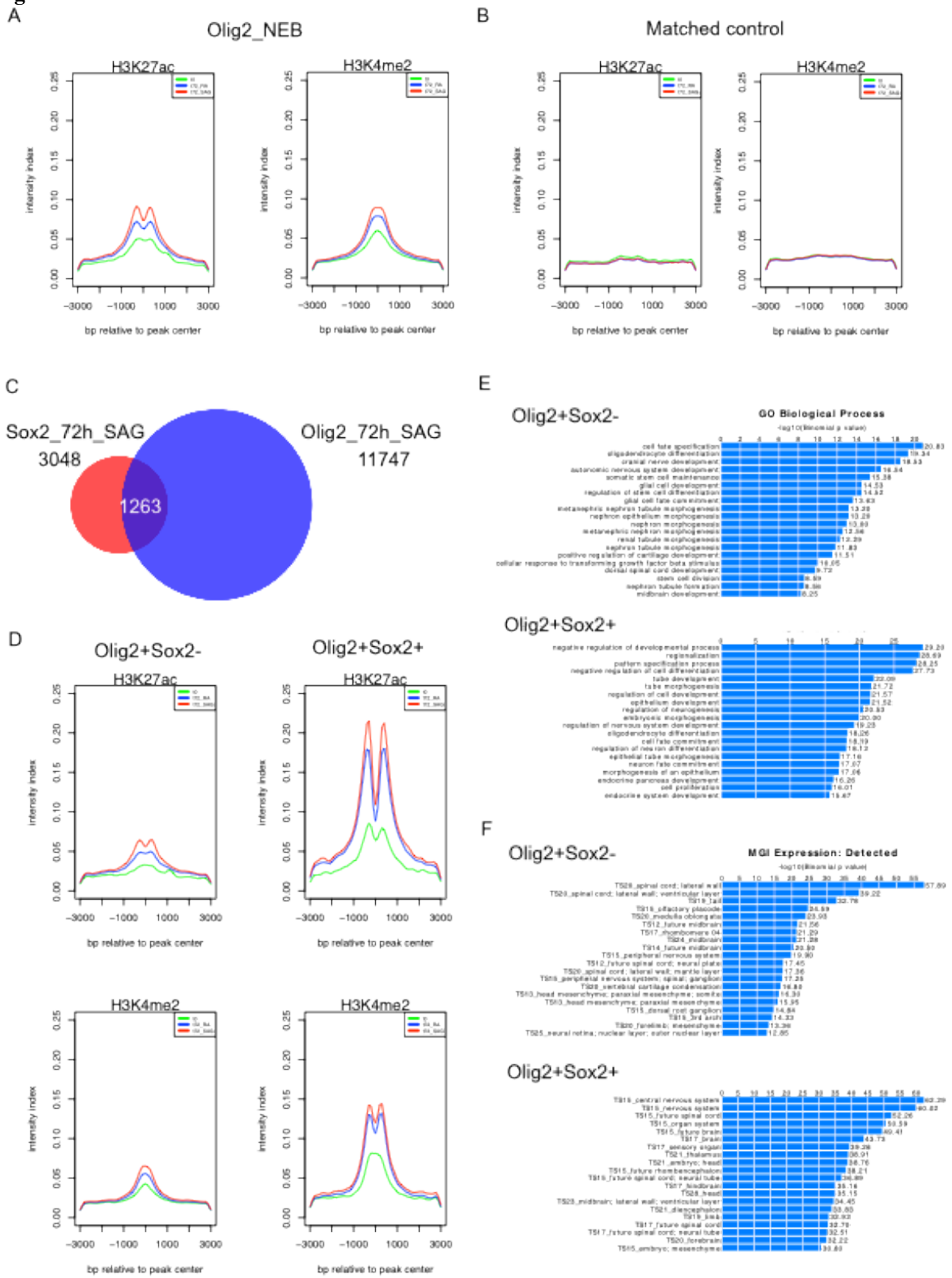

A
Olig2_NEB
H3K27ac
H3K4me2
intensity index
bp relative to peak center
B
Matched control
C
Sox2_72h_SAG
3048
1263
Olig2_72h_SAG
11747
D
Olig2+Sox2-
Olig2+Sox2+
E
GO Biological Process
F
MGI Expression: Detected

Figure 3.4 (Continued)
Active enhancer signature characterizing Olig2 NEB binding regions.
(A) Olig2 binding regions in NEB show the bimodal profile of H3K27ac and an enrichment of H3K4me2 during in vitro differentiation. Shh responsiveness is also demonstrated by higher active enhancer signature under SAG induction.
(B) Matched control regions fail to show such enrichment.
(C) Venn diagram intersection of Olig2 with Sox2 binding regions show that a significant part of the active input might come from Sox2.
(D) Olig2+Sox2+ regions indeed have stronger active enhancer signature. Left: Olig2+Sox2- regions; right: Olig2+Sox2+ regions.
(E) GREAT analysis of different types of Olig2 peaks showing that Olig2+Sox2+ co-binding peaks regulate genes earlier in development than targets of Olig2+Sox2- type of peaks.

earlier stage than Olig2+Sox2- target genes, with the latter regulating genes more differentiated in a specific neuronal lineage. Thus in identifying TFs that promote pMN transition into MN and OLP cell types, although Sox2 provides a significant active input to a portion of Olig2 binding peaks, it is likely to serve an earlier stage rather than directing a more differentiated cell fate, and there are other sources of active inputs that specifies neuronal subtype identities. For the different scenarios of Sox2 and Olig2 binding, see Figure 3.5. Other SoxB1 family members that function at a later stage, such as Sox21, promotes neurogenesis as a repressor (Sandberg et al., 2005), and is thus not our candidate.

E-box motif variants are differentially preferred by Olig2 in NEB

In searching for TFs that contribute active input to promote pMN differentiation in Olig2 *cis*-regulatory elements, I utilized de novo motif discovery and TF DNA recognition site specificity to identify potential candidates. Recent establishment of various motif databases, such as TFANSFAC, UniProbe, and Jaspar have greatly enhanced our knowledge of DNA recognition motifs for various DBD families. Meanwhile, the identification of collaborating TFs based on motif analyses is highlighted in various studies (He et al., 2010; Mullen et al., 2011; Trompouki et al., 2011). The assumption is that motif analysis in Olig2 peak regions will unveil potential collaborating TFs that help promote pMN differentiation, and different Olig2 co-factors might be associated with MN or OLP cell fates.

bHLH family factors recognize a CANNTG E-box motif and bind to DNA via heterodimers or homodimers. DNA binding motif preferences have been shown to be distinct for different bHLH factors and dimer species (Cao et al., 2010; Fong et al., 2012). In order to explore the Olig2-mediated dimer species that promote pMN differentiation to MN and OLP, we

Figure 3.5

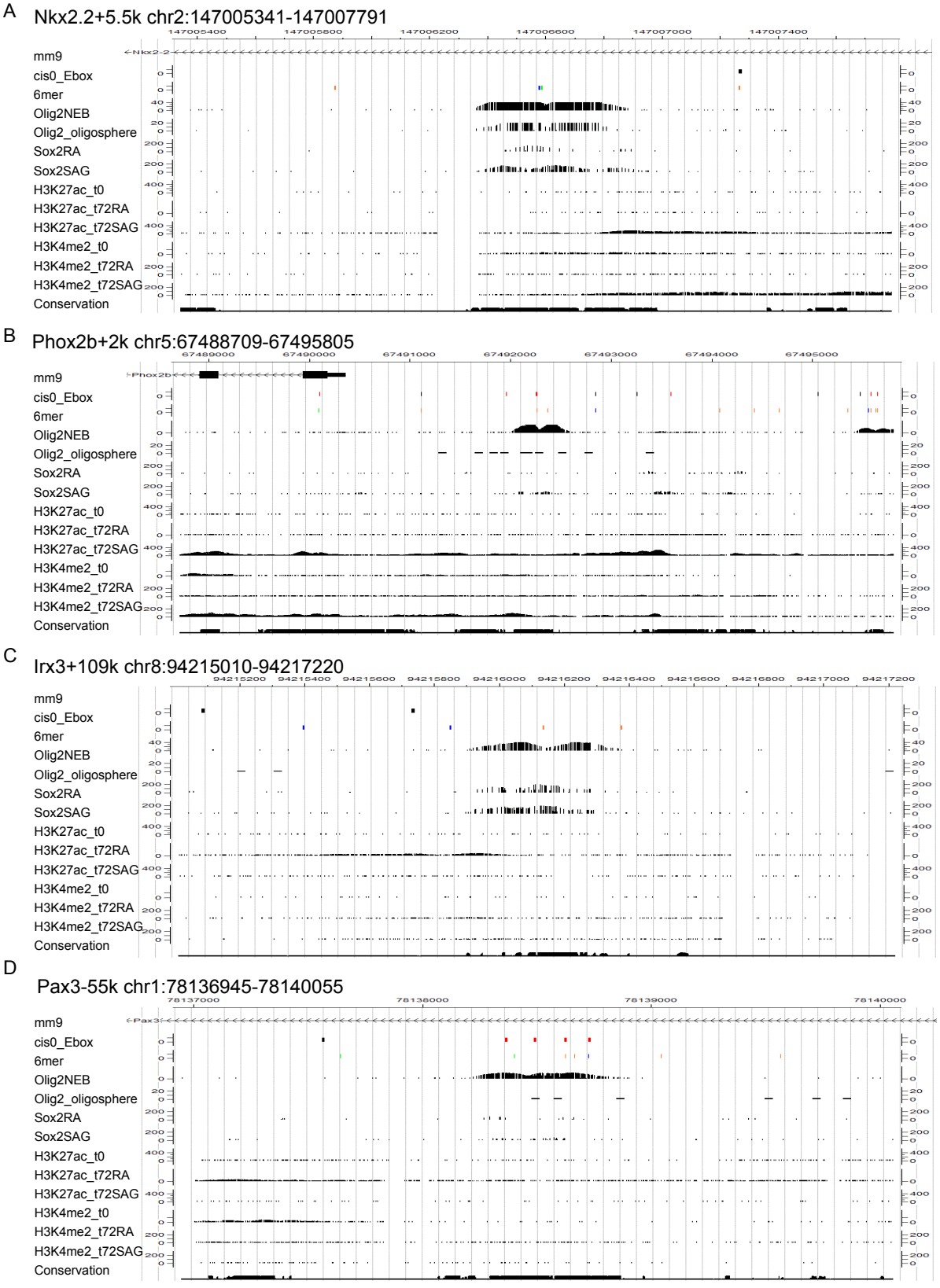

A
Nkx2.2+5.5k chr2:147005341-147007791
147005400
147005800
147006200
147006600
147007000
147007400
Nkx2.2
mm9
cis0_Ebox
6mer
Olig2NEB
Olig2_oligosphere
Sox2RA
Sox2SAG
H3K27ac_t0
H3K27ac_t72RA
H3K27ac_t72SAG
H3K4me2_t0
H3K4me2_t72RA
H3K4me2_t72SAG
Conservation
B
Phox2b+2k chr5:67488709-67495805
67489000
67490000
67491000
67492000
67493000
67494000
67495000
Phox2b
mm9
cis0_Ebox
6mer
Olig2NEB
Olig2_oligosphere
Sox2RA
Sox2SAG
H3K27ac_t0
H3K27ac_t72RA
H3K27ac_t72SAG
H3K4me2_t0
H3K4me2_t72RA
H3K4me2_t72SAG
Conservation
C
Irx3+109k chr8:94215010-94217220
94215200
94215400
94215600
94215800
94216000
94216200
94216400
94216600
94216800
94217000
94217200
mm9
cis0_Ebox
6mer
Olig2NEB
Olig2_oligosphere
Sox2RA
Sox2SAG
H3K27ac_t0
H3K27ac_t72RA
H3K27ac_t72SAG
H3K4me2_t0
H3K4me2_t72RA
H3K4me2_t72SAG
Conservation
D
Pax3-55k chr1:78136945-78140055
78137000
78138000
78139000
78140000
Pax3
mm9
cis0_Ebox
6mer
Olig2NEB
Olig2_oligosphere
Sox2RA
Sox2SAG
H3K27ac_t0
H3K27ac_t72RA
H3K27ac_t72SAG
H3K4me2_t0
H3K4me2_t72RA
H3K4me2_t72SAG
Conservation

Figure 3.5 (Continued)
CisGenome browser examples showing the combinatorial binding patterns of Olig2 and Sox2.
(A) Active enhancer signature of Olig2 binding with Sox2 binding.
(B) Active enhancer signature of Olig2 without Sox2 binding.
(C) Closed chromatin signature of Olig2 with Sox2 binding.
(D) Closed chromatin signature of Olig2 without Sox2 binding.

set out to investigate the E-box 6mer motif that is preferred by Olig2 using ChIP-seq data and protein binding microarray (PBM). ChIP-seq will reflect Olig2 sequence preference in physiological state, while PBM will be *in vitro* biochemical preference site for Olig2 homodimer only.

In NEB, we noticed an enrichment of some of the ten theoretically possible 6mer variants in comparison when compared with those predicted in the genome (Figure 3.6A). In particular, CAGCTG and especially CATCTG show the strongest enrichment, 1.33 fold and 2.57 fold, respectively, over genomic background. If we plot distribution of the different motif variants around Olig2 peak center in NEB, we further discovered CATATG motif being centered near Olig2 peak center, in addition to CAGCTG and CATCTG (Figure 3.6B). The centered distribution of motif variants near Olig2 peak center is strong evidence that these DNA sequences are preferably bound by Olig2 in a non-random manner. Interestingly, the PBM recovered Olig2 motif is significantly different from that recovered from ChIP-seq (Figure 3.6C): the PBM motif has a consensus CATATG, while ChIP-seq motif has a consensus CAT(G)C(A)TG. Considering the dyad-symmetry of the former versus latter motif, and that Olig2 homodimers bind the former in PBM studies, the collective data suggest the ChIP-seq data is a mix of homo- and heterodimeric Olig2 containing bHLH complexes.

We separated Olig2 NEB peaks depending on the E-box motif variants' presence within 50bp from peak center. CATATG-containing peaks have a higher rank out of all Olig2 peaks, pointing to a potential higher affinity of Olig2 dimer to CATATG 6mer compared to other protein-motif interactions (Figure 3.6D). Interestingly, if we explore the average RNA-seq readout FPKM values for genes near peaks with different 6mer variants, we find genes near CATATG containing peaks have lower expression levels on average (Figure 3.6E), suggesting

Figure 3.6

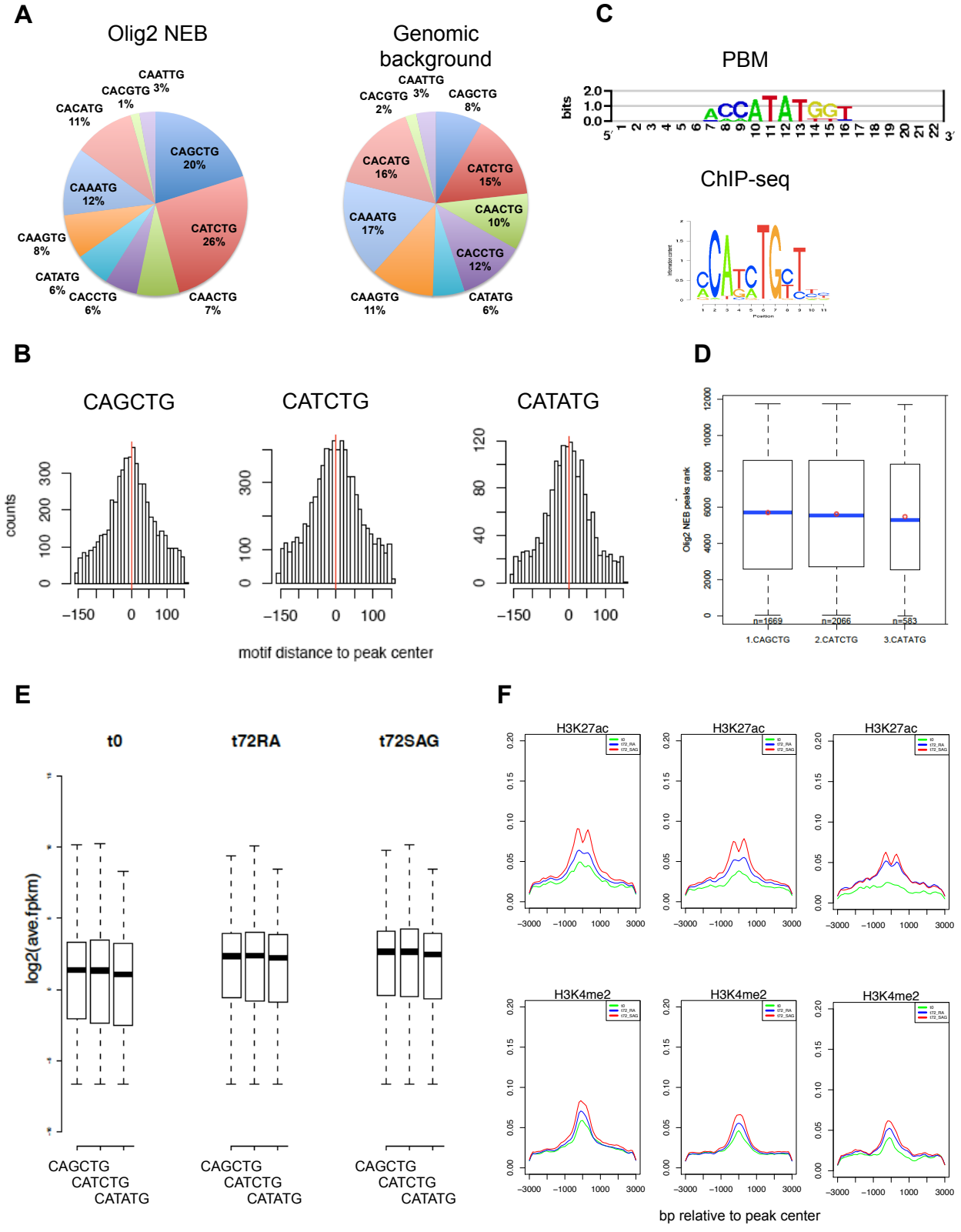
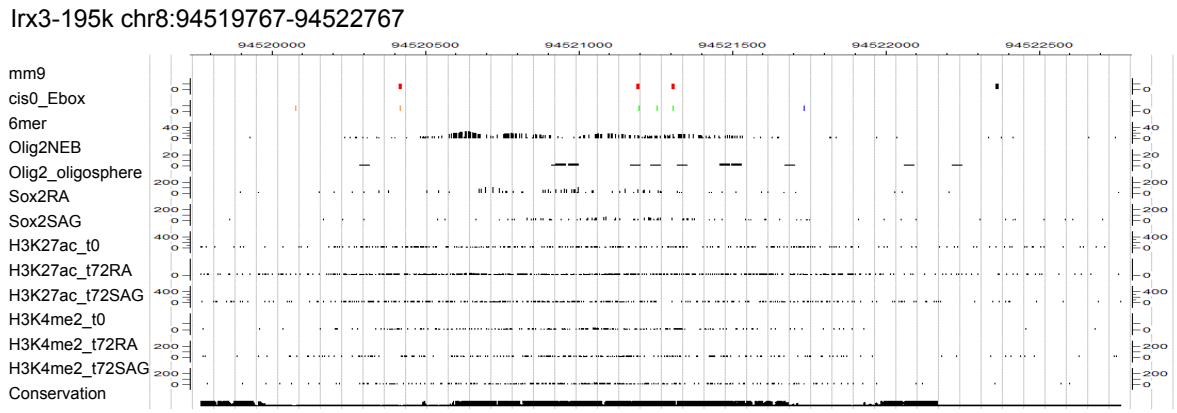
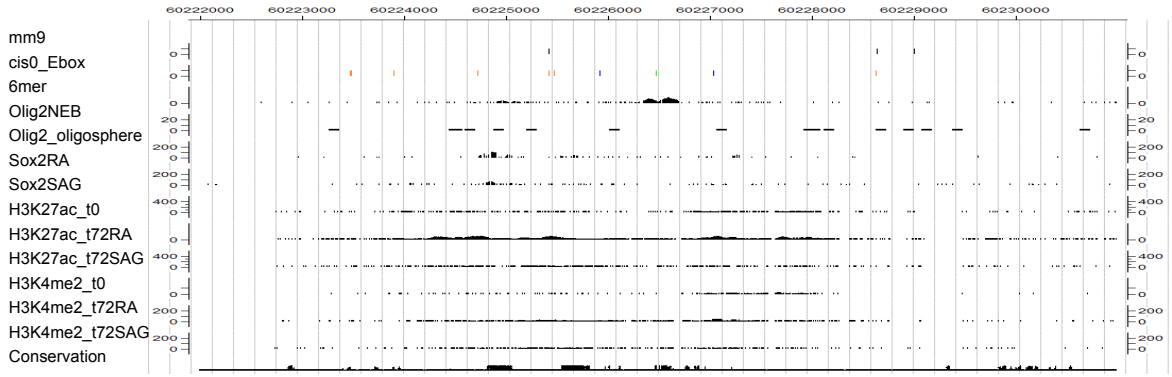

A
Olig2 NEB
CAGCTG 20%
CATCTG 26%
CAACTG 7%
CACCTG 6%
CATATG 6%
CAAGTG 8%
CAAATG 12%
CACATG 11%
CACGTG 1%
CAATTG 3%
Genomic background
CAGCTG 8%
CATCTG 15%
CAACTG 10%
CACCTG 12%
CATATG 6%
CAAGTG 11%
CAAATG 17%
CACATG 16%
CACGTG 2%
CAATTG 3%
C
PBM
bits
ChIP-seq
B
CAGCTG
CATCTG
CATATG
counts
motif distance to peak center
D
Olig2 NEB peaks rank
n=1669
n=2066
n=583
1.CAGCTG
2.CATCTG
3.CATATG
E
t0
t72RA
t72SAG
log2(ave.fpkm)
CAGCTG
CATCTG
CATATG
F
H3K27ac
H3K4me2
bp relative to peak center

Figure 3.6 Continued

G



H



I

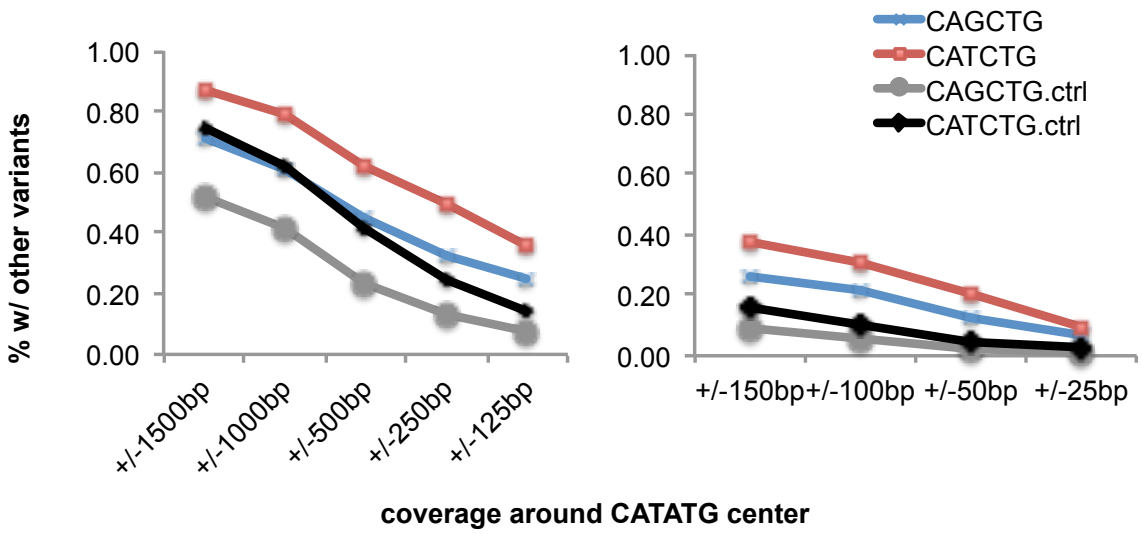


Figure 3.6 (Continued)
E-box motif variants are differentially preferred by Olig2 in NEB.

(A) E-box 6mer variants occurrences in Olig2 NEB binding regions in comparison with genomic background frequency. Left: Olig2 NEB binding regions; right: genomic background.

(B) Frequency of three variants that show centered distribution relative to Olig2 peak center.

(C) E-box motif recovered from protein binding microarray (PBM, top) and ChIP-seq in NEB (ChIP-seq, bottom) show differential preference to the CANNTG core 6mer with regards to the center two nucleotide.

(D) Rank of Olig2 NEB peaks with different E-box variants within 25bp from peak center. CATATG-containing peaks show smaller ranks (ie, stronger peak signals) than CATCTG and CAGCTG peaks, suggesting higher affinity of Olig2 dimer to CATATG motif in comparison with the other two variants. Red: mean of ranks; blue: median of ranks.

(E) Boxplot showing low expression levels for gene regulated by CATATG-mediated Olig2 binding, in comparison with CAGCTG and CATCTG-mediated Olig2 binding peaks.

(F) Active enhancer signature H3K27ac and H3K4me2 near Olig2 peaks containing CAGCTG (left), CATCTG (middle), and CATATG (right) E-box motif variants.

(G)-(H) Genome browser showing lack of active enhancer signature for Olig2 binding sites mediated by CATATG. (G) Cis-element 195kb downstream from Irx3 TSS. (H) Cis-element 55 kb upstream from Gas1 TSS. For the '6mer' track showing different E-box variant: blue: CAGCTG; orange: CATCTG; green: CATATG.

(I) Percentage of CATATG Olig2 peaks with CATCTG or CAGCTG consensus within a certain distance from CATATG motif center, in comparison with CATATG motifs randomly sampled. Window coverage is from +/-1500 to +/-125bp (left) and from +/-150bp to +/-25bp (right.)

the homodimer binding has a signature consistent with repression. Further, features of active enhancer signatures were less enriched as well as Shh-responsiveness when CATATG containing peaks were compared to CAGCTG and CATCTG peaks (Figure 3.6F). Examples of peaks with CATATG variant are shown in Figure 3.6G-H.

We also notice a tendency for the occurrence of multiple E-box sequences when we observe CATATG variant. Olig2 peaks with the CATATG variant +/-25bp from an Olig2 ChIP-peak center were pooled, and the percentage of these peaks with additional CAGCTG or CATCTG variants within different windows were calculated (Figure 3.6I). If the clustering of E-box sequences are a genome-wide phenomenon, we expect to see higher % of CATATG Olig2 peaks to have CAGCTG or CATCTG motifs nearby than control CATATG sequences not within Olig2 peak centers.

We observed a clear tendency of more CAGCTG or CATCTG motifs occurring in Olig2 peaks with a centered CATATG variant: the red and green lines above the black and grey lines in Figure 3.6I. The enrichment over control increases as the window decreases from +/-1500bp to +/-125bp suggesting a non-random local clustering (Figure 3.6I left). As the enrichment decreases as the window decreases below +/-100bp the data do not suggest a tightly constrained clustering reflective of cooperative interactions (Fig. 3.6I right). Together the result suggest that multiple motifs likely represent independent protein binding events, and having multiple motifs might add the robustness of the regulatory process.

Olig2 peaks with different E-box variants are associated with genes of distinct expression patterns and biological processes

Analyses so far suggest that Olig2 peaks with different E-box variants have distinct histone modification profiles, and likely regulate genes with differential expression levels. A recent study provided a mechanism whereby distinct differentiation can be achieved by differential E-box sequence preferences mediated by bHLH lineage-specific factors (Fong et al., 2012). Specifically, Fong et al. (2012) profiled chromatin binding of two E-box TFs, NeuroD2 and MyoD, in two contexts: in the P19 cell line transduced with NeuroD2 and converted to neurons, and mouse embryonic fibroblast (MEF) transduced with MyoD and converted to skeletal muscle cells. They identified common and distinct binding regions for NeuroD2 and MyoD. Interestingly, only the distinct binding sites are associated with lineage specific differentiation, while the common peaks are not strongly associated with gene transcription regulation.

This result gives several insights. First, it supports the concept of utilizing motif preferences to study cell fate choices to different neuronal subtypes. The hypothesis is that the generation of mitotically active MN and OLP precursors might be achieved by different Olig2 dimers with different sequence specificity preferences. Second, NeuroD2 and MyoD genome-wide chromatin binding profiles provide a valuable information source to help identify Olig2 dimerization partners in our data. Thus, I integrated Olig2 NEB ChIP-seq with that of MyoD and NeuroD2 (Figure 3.7A).

The first question I asked is whether a certain E-box motif variant is lineage specific, while others may be a generic E-box consensus. To this end, I investigated enriched E-box 6-mer variants in Olig2, NeuroD2, and MyoD ChIP-seq datasets. CAGCTG and CATCTG motifs were reported to be enriched for NeuroD2 (Fong et al., 2012). The CATATG variant is also strongly

Figure 3.7

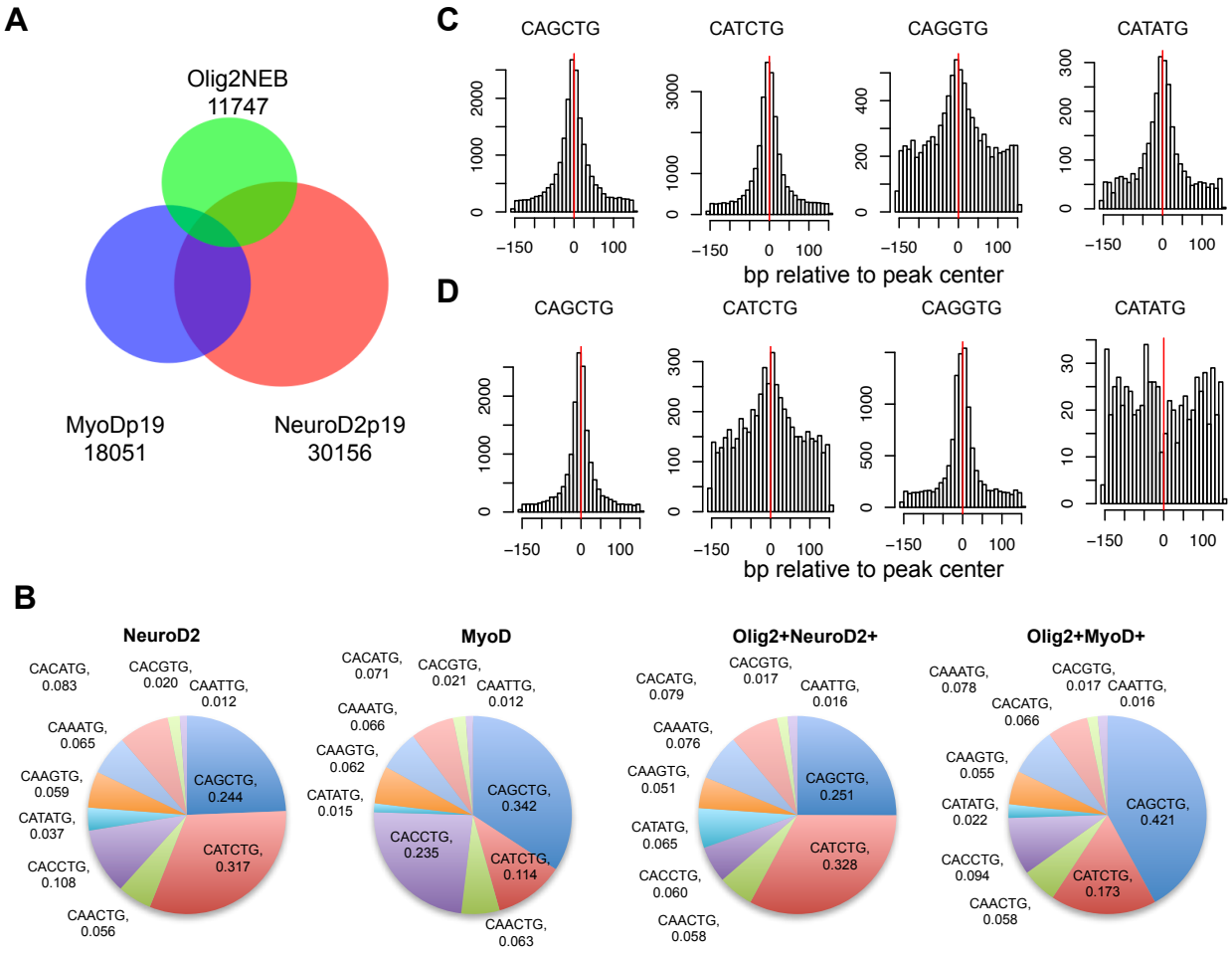

A
Olig2NEB
11747
MyoDp19
18051
NeuroD2p19
30156
C
CAGCTG
CATCTG
CAGGTG
CATATG
bp relative to peak center
D
CAGCTG
CATCTG
CAGGTG
CATATG
bp relative to peak center
B
NeuroD2
CACATG, 0.083
CACGTG, 0.020
CAATTG, 0.012
CAAATG, 0.065
CAAGTG, 0.059
CATATG, 0.037
CACCTG, 0.108
CAACTG, 0.056
CAGCTG, 0.244
CATCTG, 0.317
MyoD
CACATG, 0.071
CACGTG, 0.021
CAATTG, 0.012
CAAATG, 0.066
CAAGTG, 0.062
CATATG, 0.015
CACCTG, 0.235
CAACTG, 0.063
CAGCTG, 0.342
CATCTG, 0.114
Olig2+NeuroD2+
CACATG, 0.079
CACGTG, 0.017
CAATTG, 0.016
CAAATG, 0.076
CAAGTG, 0.051
CATATG, 0.065
CACCTG, 0.060
CAACTG, 0.058
CAGCTG, 0.251
CATCTG, 0.328
Olig2+MyoD+
CAAATG, 0.078
CACGTG, 0.017
CAATTG, 0.016
CACATG, 0.066
CAAGTG, 0.055
CATATG, 0.022
CACCTG, 0.094
CAACTG, 0.058
CAGCTG, 0.421
CATCTG, 0.173

Figure 3.7 Continued

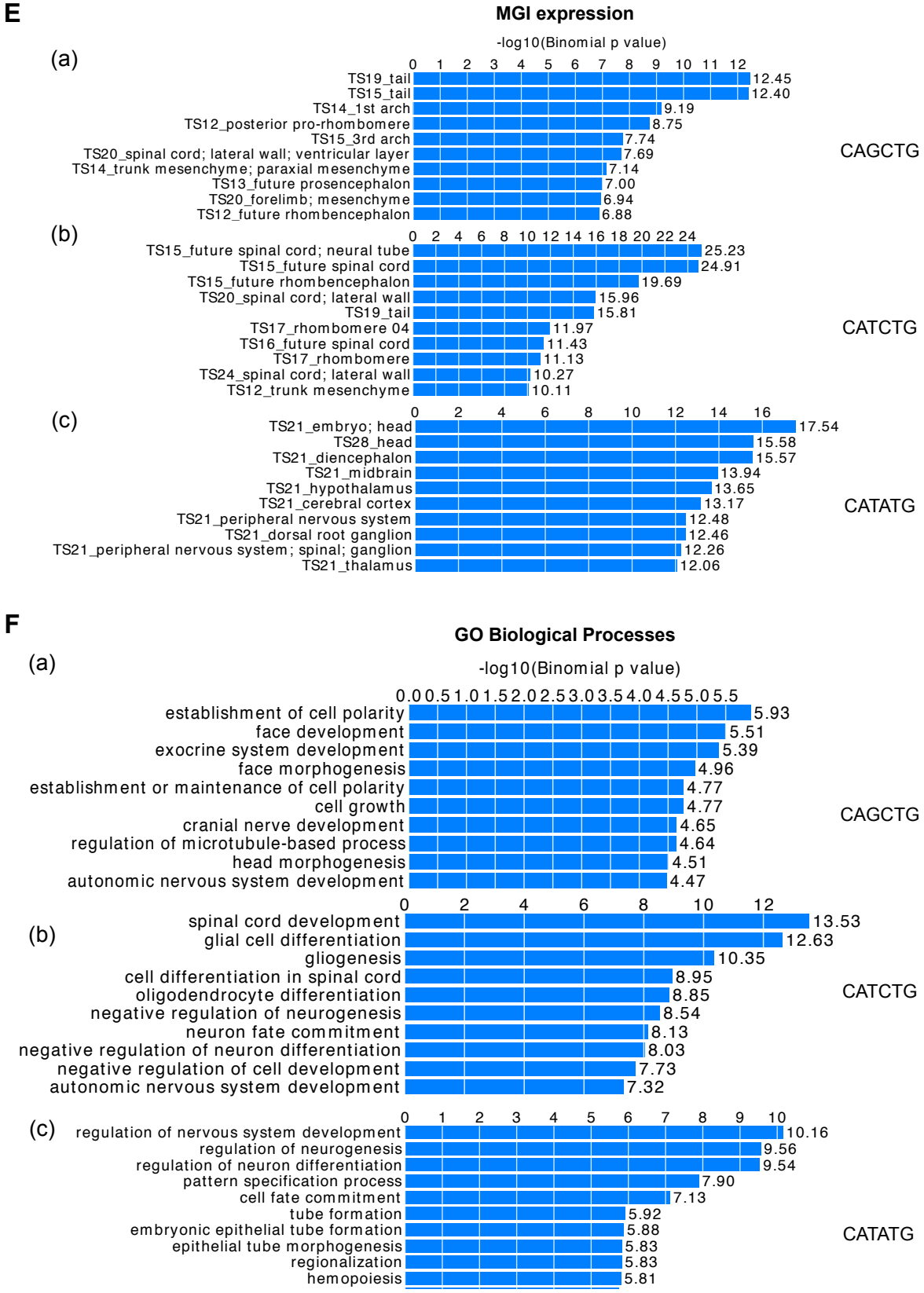


Figure 3.7 (Continued)
Olig2 peaks with different E-box variants are associated with genes with distinct expression domains and involved in different biological processes.
(A) 3-way intersection of Olig2, MyoD and NeuroD2 ChIP-seq data.
(B) E-box 6mer variants occurrences in NeuroD2 and MyoD peak regions, as well as common peaks between NeuroD2 and Olig2 as well as MyoD and Olig2.
(C) –(D) Frequency of representative E-box variants relative to NeuroD2 (C) and MyoD (D) peak center.
(E) Olig2 peaks with different 6mer variants show different MGI expression stage and tissues. (a): CAGCTG Olig2 peaks; (b): CATCTG Olig2 peaks; (c): CATATG Olig2 peaks.
(F) Olig2 peaks with different 6mer variants show different GO enrichment. (a): CAGCTG Olig2 peaks; (b): CATCTG Olig2 peaks; (c): CATATG Olig2 peaks.

centered around NeuroD2 peak center, and to a less extent CAGGTG (Figure 3.7B-a). In contrast, in the MyoD ChIP-seq, a CATATG motif is depleted (Figure 3.7B-b) and centering is not apparent (Figure 3.7D). Recall that CATATG encodes the optimal Olig2 homodimer motif recovered from PBM (Figure 3.6C top). This is the sequence strongly centered in the neural Olig2 ChIP-seq peaks (Figure 3.6B right). Thus, this motif correlates with neuronal but not muscle datasets. Meanwhile, if we expect certain E-box variant to be predictive of lineage differentiation and some to be only generic variants, then the common binding peaks between different E-box TFs should have an enrichment of the generic variants, while the distinct peaks will be enriched for lineage specific variants. If we do 2-way intersection of Olig2 with MyoD and examine the motif variants constitution (Figure 3.7B-d), we observe an increase of CAGCTG motif to 42% of all E-box variants for the shared Olig2+MyoD+ peak regions, suggesting that CAGCTG sites might indeed represent a more generally utilized recognition site for various E-box TFs independent of cell origin. In contrast, similar analysis on Olig2 and NeuroD2 intersection gave a different result; both CAGCTG and CATCTG motifs are consistently enriched (Figure 3.7D-c), but more especially CATCTG (Figure 3.6A right). I conclude that CAGCTG sequence may be commonly recognized by E-box proteins of different lineages. In contrast, CATCTG and CATATG motifs are likely more significant within neuronal lineage specific differentiation.

Next, I asked whether Olig2-mediated differentiation to MN and OLP lineages are characterized by different E-box 6-mer variant preferences that might reflect Olig2 dimer partner preferences. GO analysis was performed on predicted target genes of different Olig2 dimer partnerships. In terms of MGI expression patterns (Figure 3.7E), CAGCTG Olig2 peak regulated genes have a broad spectrum of tissue expression, not restricted to spinal cord and neuronal

lineages. In contrast, CATCTG peak regulated genes are very specifically NT-related genes with high enrichment values. Interestingly, through GO analysis (Figure 3.7F), I discovered that glial differentiation-related terms, in particular oligodendrocyte differentiation, are enriched for CATCTG but not CATATG associated Olig2 peaks (Figure 3.7F b,c). These observations suggest: 1) CATCTG is the core variant recognized by Olig2 in spinal cord motor neuron progenitors and may also be important for later glial cell differentiation into OLP; 2) CATATG might function in MN development at a later stage but has little connection to glial lineage; 3) CAGCTG is more broadly utilized by bHLH members in non-tissue specific contexts.

Discussion

Our transcriptional analysis of Shh-mediated embryonic neural patterning has generated several new insights into the GRNs by elucidating several layers of information that relate TF occupancy to chromatin accessibility to gene expression regulation. We propose that as the first stage of initiating Shh response, Sox2 functions to create a permissive chromatin structure for Gli1 access; as development proceeds, ventral repressive determinants interact with co-activators to promote further differentiation. In addition, different Olig2-mediated dimer species is one potential mechanism for Olig2-mediated cell fate choice from pMN to MN or OLPs.

Shh neural-specific response by integration of Gli1-defined neural CRMs with Sox2 input

Developmental studies have highlighted a small number of signaling pathways mediating a great diversity of tissue patterning, raising the question of how tissue-specific outcomes arise from common signaling inputs. ChIP-seq integration of Gli1 and Sox2 established Sox2 as a necessary activating input for Gli1-mediated Shh response by having an accessible chromatin structure. We demonstrated that Gli1 acts in conjunction with Sox activity to switch on distinct ventral neural progenitor determinants. Shh regulation of Gli binding provides the instructive input to direct specific ventral cell fates, and Sox binding provides the neural context. Genetic evidence indicates that SoxB1 members regulate the proliferation and timing of differentiation of neural progenitors, and recent evidence also links their actions to diversification of the progenitor pool (Bergsland et al. 2011). In a recent model, Sox factors are proposed to bind sequentially to neural target genes, with replacement of Sox2 in ESCs followed by Sox3 and then finally Sox11 in mature neurons (Bergsland et al. 2011). Although several neural-specific genes are associated with Sox2-bound regions in ESC studies, the putative CRMs highlighted by Sox2 binding do not

display a permissive chromatin mark until sometime later, on commitment to a specific neural lineage, suggesting that Sox2 engagement is not sufficient to modify the chromatin signature (Bergsland et al. 2011).

To address whether Sox2 primes Class-II ventral gene activation preceding Gli factor engagement, will require a more comprehensive time-course examining chromatin binding patterns of Sox2 and Gli1 using ChIP-qPCR or ChIP-seq. The expectation is the precedence of Sox2 binding relative to Gli1, which might act to open up chromatin to create an assessable structure. However, all data to date is based on ectopic activities. Incorporating endogenous actions of all Gli factors is needed, and the study on Gli1 may only reveal a certain set of information. For example, the presence of another Gli factor in the VNT, Gli2, and current data suggesting the up-regulation of Gli2 as early as 12 hours into induction might complicate the interpretation of any Sox2N prior to Gli1 binding. Meanwhile, since only 12% of Gli target genes are also bound by Sox2, it remains to be studied whether there is other alternative mechanism for Gli-mediated neural-specific responses, in addition to the proposed Sox2 priming model.

Common cis-regulatory elements for VNT repressor TFs

The extensive overlap of binding among ventral determinants Nkx2.2, Nkx6.1 and Olig2 within what is likely the same CRM further suggested the clear delineation of the adjacent progenitor cell types via cross-repression and enabled the systematic identification of *cis*-regulatory elements in neural derivatives. GO analysis of the co-regulated genes by the ventral repressors showed several major classes of target genes are enriched: Class-II ventral cell determinants, Class-I dorsal determinants, post-mitotic neuronal subtype determinants, and Hh pathway

components. Post-mitotic TFs are particular interesting. This strong enrichment of post-mitotic neuronal fate determinants by repressor binding sites is a tantalizing connection between terminal differentiation and repressor factors. In understanding this gene regulation, identifying activating inputs functioning with these repressors is critical. Regionalized versus ubiquitous activator input will make a significant difference in understanding the gene regulatory principles by these repressors.

bHLH TF DNA recognition site preferences for MN versus OLP fate choices

We provided evidence supporting that differential DNA recognition site preferences are dependent on different co-factors. Olig2 DNA binding differential preferences in MNs and OLPs, together with gene expression level information, points to a model of Olig2::X heterodimer promoting OLP cell fate by preferentially binding to CATCTG sequences, in contrast to the Olig2::Olig2 homodimer preferences of CATATG in MNs. The next step will be the prediction of Olig2 heterodimer partners in MN and OLP lineages. Based on RNA-seq expression data in our system and existing expression data from in vivo isolated cells in neuron, oligodendrocyte, and astrocyte lineages (Cahoy et al., 2008), NeuroD1 and Ascl1 are the most likely bHLH factor candidate for Olig2 dimer, given their strong expression in NEB, and high fold enrichment in neuron (16 folds) and OLPs (14 folds). Neurog2 and NeuroD4 are also likely candidates for MN fate according to prior knowledge, but their expression enrichment data are not available from Cahoy et al. (2008) and further experiments need to confirm.

A more complete understanding of Olig2 complexes and their actions would benefit from a number of approaches. First, PBM can be utilized to study DNA sequence specificity for the proposed Olig2 dimers by forced tethering of Olig2 with candidates dimer partners, although

steric hindrance on forced tethering, and the absence of helper could influence these results. Second, ectopic expression of Olig2 dimers in the chick NT enables their activities to be assayed. For example, based on our hypothesis, the ectopic expression of Olig2::Ascl1 heterodimer should contribute to ectopic generation of OLPs. Third, in order to validate activator versus repressor function of a certain dimer species, luciferase assay can be performed once the DNA binding specificities for different dimer species are confirmed. Finally, the in vivo chromatin binding specificity of different Olig2 dimer species will be verified by ChIP against tethered Olig2 dimers in neural EB. The neural EB will be differentiated from mESCs expressing tethered dimers, which can be generated using recombinase-mediated cassette exchange (RMCE) in mESCs. In the cases that certain antibodies are not available, RMCE system makes the usage of epitope tags possible.

Materials and methods

Neural progenitor culture

ES cells were maintained on mouse embryonic fibroblast (MEF) feeder cells and cultured in DMEM with 15-17.5% FBS supplemented with LIF. To generate neural progenitors from ES cells, we followed the NEB culture protocol described in (Wichterle et al., 2002). First, MEFs were removed in a gelatin-coated flask and floating ES cells were collected. Cells were rinsed once with DFNB medium (DMEM: F12: Neurobasal=1:1:2, 4% B27 supplement; Gibco) or DFNK medium (10% Knockout Serum Replacement in place of B27). Cells were plated in Ultra Low Attachment 6-well plates or T75 flask at 100,000-200,000/ml density in DFNB/DFNK. Two days after MEF depletion, the medium was changed to a fresh DFNB/DFNK medium

containing 500nM all-trans RA (Sigma) and 50-800nM SAG (Alexis Biochemicals). Embryoid bodies were harvested 72hrs post induction unless otherwise noted.

Microarray analysis

Embryoid bodies were generated with the above protocol. Two days after MEF depletion, EBs were treated with 500nM RA or 500nM RA and 50nM SAG for 3 days. Whole fraction RNA was extracted with mirVana RNA kit (Ambion) according to the manufacturer's instruction. Three pairs of biological replicates were collected. RNA samples were validated with Bioanalyzer (Agilent) and biotin-labeled cRNA was prepared with 3' IVT Express kit (Affymetrix). The samples were and hybridized to Mouse 430 2.0 arrays (Affymetrix).

ChIP-seq data analysis

The first 25-nt of each read was mapped to mouse mm9 genome assembly using SeqMap (Jiang and Wong, 2008) or ELAND (Illumina). Two-sample peak detection was performed using an iterative conditional binomial model (Ma and Wong, 2011). For one-sample peak calling, the FDR of each w-bp window with k_i reads is the ratio of expected number of windows containing at least k_i reads (given the negative binomial parameters) to the observed number of windows that have at least k_i reads. In the two-sample peak detection, the FDR of each w-bp window with k_{1i} ChIP reads and n_i total reads, is the ratio of the expected number of windows (given r0) to the observed number of windows that have at least k_{1i} ChIP reads out of a n_i-read window (see Methods in (Ji et al., 2008)). In the peak calling process, the forward and reverse strand reads were processed separately. First, peaks were identified for each stand and then forward and reverse strand peaks were combined if they were within 500 bp of each other and the

corresponding number of reads in the coupled peaks shows less than five-fold change. The start and end coordinates for peak boundaries were defined as the modes of the coupled peaks. A final peak score was then determined by averaging the heights of the forward and reverse-called peaks. The iterative two-sample peak calling was performed separately on the Gli1-A and the Gli1-B replicates using the YFP mock input control with FDRs: 0.01, 0.05 and 0.1. Additional ChIP-set data sets were obtained for: Sox2 (ESC) (Marson et al., 2008), H3K4me2 (ESC) (Meissner et al., 2008) and H3K27ac (ESC) (Creyghton et al., 2010). Sox2-ESC peak calls were performed as above.

Gene and enhancer annotation

Gene ontology analyses were performed using DAVID (http://david.abcc.ncifcrf.gov/) (Huang et al., 2009). The association of genome regions with gene ontology terms was done using GREAT (version 1.8.2) (http://great.stanford.edu/) (Bejerano et al., 2010).

Histone modification analysis

Aggregates plot of histone modifications were performed using Homer (Heinz et al., 2010). Genomic regions upstream and downstream of a certain distance from the peak center were divided into 60 100-bp bins each. Error bars indicate standard errors of the means of all the normalized ChIP-seq fragment densities in the respective bin. Box plot of expression values was drawn in R. Statistical significance was calculated by the Wilcoxon rank-sum test.

Hierarchical clustering, heat map and Venn diagram

For the hierarchical clustering of 20 factors based on chromatin co-occupancy, peak regions for each factor with p-value < 1e-8 were trimmed or expanded to 400 bp centered at the peak center. Pair-wise intersection was performed between every pair of factors to generate a matrix of the number of co-bound regions. Complete linkage, correlation distance, hierarchical clustering was performed and heat map generated using R. Peak-peak intersection were conducted using BEDTools (v2.10.1) (Quinlan and Hall, 2010). Venn diagrams were draw in Cistrome analysis pipeline(Liu et al., 2011).

RNA-seq analysis

RNA-seq reads were mapped using TopHat with default settings (http://tophat.cbcb.umd.edu) (Roberts et al., 2012). Tophat output data were analyzed by Cufflinks to generate FPKM values for both known and de novo transcripts in mouse genome. Cuffdiff was then used to calculated significant differential expression across conditions.

References

Anderson, D.J. (1995). Neural Development - Spinning Skin into Neurons. Curr Biol *5*, 1235-1238.

Barski, A., Cuddapah, S., Cui, K.R., Roh, T.Y., Schones, D.E., Wang, Z.B., Wei, G., Chepelev, I., and Zhao, K.J. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. Gene Dev *25*, 2453-2464.

Bylund, M., Andersson, E., Novitch, B.G., and Muhr, J. (2003). Vertebrate neurogenesis is counteracted by Sox1–3 activity. Nat Neurosci *6*, 1162-1168.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A.*, et al.* (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. J Neurosci *28*, 264-278.

Cao, Y., Yao, Z.Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., MacQuarrie, K.L., Davison, J., Morgan, M.T., Ruzzo, W.L.*, et al.* (2010). Genome-wide MyoD Binding in Skeletal Muscle Cells: A Potential for Broad Cellular Reprogramming. Dev Cell *18*, 662-674.

Castro, D.S., Skowronska-Krawczyk, D., Armant, O., Donaldson, I.J., Parras, C., Hunt, C., Critchley, J.A., Nguyen, L., Gossler, A., Gottgens, B.*, et al.* (2006). Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. Dev Cell *11*, 831-844.

Chen, J.K., Taipale, J., Young, K.E., Maiti, T., and Beachy, P.A. (2002). Small molecule modulation of Smoothened activity. P Natl Acad Sci USA *99*, 14071-14076.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. P Natl Acad Sci USA *107*, 21931-21936.

Fong, A.P., Yao, Z.Z., Zhong, J.W., Cao, Y., Ruzzo, W.L., Gentleman, R.C., and Tapscott, S.J. (2012). Genetic and Epigenetic Determinants of Neurogenesis and Myogenesis. Dev Cell *22*, 721-735.

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M.*, et al.* (2010). Nucleosome dynamics define transcriptional enhancers. Nat Genet *42*, 343-347.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell *38*, 576-589.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols *4*, 44-57.

Ingham, P.W., and McMahon, A.P. (2001). Hedgehog signaling in animal development: paradigms and principles. Gene Dev *15*, 3059-3087.

Ji, H.K., Jiang, H., Ma, W.X., Johnson, D.S., Myers, R.M., and Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat Biotechnol *26*, 1293-1300.

Kageyama, R., and Nakanishi, S. (1997). Helix-loop-helix factors in growth and differentiation of the vertebrate nervous system. Curr Opin Genet Dev *7*, 659-665.

Kageyama, R., and Ohtsuka, T. (1999). The Notch-Hes pathway in mammalian neural development. Cell Res *9*, 179-188.

Kageyama, R., Ohtsuka, T., Shimojo, H., and Imayoshi, I. (2008). Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition. Nat Neurosci *11*, 1247-1251.

Kishi, M., Mizuseki, K., Sasai, N., Yamazaki, H., Shiota, K., Nakanishi, S., and Sasai, Y. (2000). Requirement of Sox2-mediated signaling for differentiation of early Xenopus neuroectoderm. Development *127*, 791-800.

Kuspert, M., Hammer, A., Bosl, M.R., and Wegner, M. (2011). Olig2 regulates Sox10 expression in oligodendrocyte precursors through an evolutionary conserved distal enhancer. Nucleic Acids Res *39*, 1280-1293.

Lee, J.E., Hollenberg, S.M., Snider, L., Turner, D.L., Lipnick, N., and Weintraub, H. (1995). Conversion of Xenopus Ectoderm into Neurons by Neurod, a Basic Helix-Loop-Helix Protein. Science *268*, 836-844.

Lee, J.K., Cho, J.H., Hwang, W.S., Lee, Y.D., Reu, D.S., and Suh-Kim, H. (2000). Expression of neuroD/BETA2 in mitotic and postmitotic neuronal cells during the development of nervous system. Dev Dynam *217*, 361-367.

Lee, S.K., Lee, B., Ruiz, E.C., and Pfaff, S.L. (2005). Olig2 and Ngn2 function in opposition to modulate gene expression in motor neuron progenitor cells. Gene Dev *19*, 282-294.

Lei, Q.B., Jeong, Y., Misra, K., Li, S.K., Zelman, A.K., Epstein, D.J., and Matise, M.P. (2006). Wnt signaling inhibitors regulate the transcriptional response to morphogenetic Shh-Gli signaling in the neural tube. Dev Cell *11*, 325-337.

Li, H.L., de Faria, J.P., Andrew, P., Nitarska, J., and Richardson, W.D. (2011). Phosphorylation Regulates OLIG2 Cofactor Choice and the Motor Neuron-Oligodendrocyte Fate Switch. Neuron *69*, 918-929.

Liu, M., Pleasure, S.J., Collins, A.E., Noebels, J.L., Naya, F.J., Tsai, M.J., and Lowenstein, D.H. (2000a). Loss of BETA2/NeuroD leads to malformation of the dentate gyrus and epilepsy. P Natl Acad Sci USA *97*, 865-870.

Liu, Q., Zhao, N.M., Yamaguch-Shinozaki, K., and Shinozaki, K. (2000b). Regulatory role of DREB transcription factors in plant drought, salt and cold tolerance. Chinese Sci Bull *45*, 970-975.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*.

Lu, Q.R., Sun, T., Zhu, Z.M., Ma, N., Garcia, M., Stiles, C.D., and Rowitch, D.H. (2002). Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection. Cell *109*, 75-86.

Ma, W.X., and Wong, W.H. (2011). The Analysis of Chip-Seq Data. Methods in Enzymology, Vol 497: Synthetic Biology, Methods for Part/Device Characterization and Chassis Engineering, Pt A *497*, 51-73.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Meissner, A., Mikkelsen, T.S., Gu, H.C., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X.L., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-U791.

Mizuguchi, R., Sugimori, M., Takebayashi, H., Kosako, H., Nagao, M., Yoshida, S., Nabeshima, Y., Shimamura, K., and Nakafuku, M. (2001). Combinatorial roles of Olig2 and Neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. Neuron *31*, 757-771.

Mullen, A.C., Orlando, D.A., Newman, J.J., Loven, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-beta Signaling. Cell *147*, 565-576.

Murre, C., Mccaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B.*, et al.* (1989). Interactions between Heterologous Helix-Loop-Helix Proteins Generate Complexes That Bind Specifically to a Common DNA-Sequence. Cell *58*, 537-544.

Nieto, M., Schuurmans, C., Britz, O., and Guillemot, F. (2001). Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. Neuron *29*, 401-413.

Novitch, B.G., Chen, A.I., and Jessell, T.M. (2001). Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. Neuron *31*, 773-789.

Peterson, K.A., Nishi, Y., Ma, W.X., Vedenko, A., Shokri, L., Zhang, X.X., McFarlane, M., Baizabal, J.M., Junker, J.P., van Oudenaarden, A.*, et al.* (2012). Neural-specific Sox2 input and differential Gli-binding affinity provide context and positional information in Shh-directed neural patterning. Gene Dev *26*, 2802-2816.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., and Trapnell, C. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

Roh, T.-Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proc Natl Acad Sci USA *103*, 15782-15787.

Sandberg, M., Kallstrom, M., and Muhr, J. (2005). Sox21 promotes the progression of vertebrate neurogenesis. Mech Develop *122*, S177-S177.

Scotting, P.J., and Rex, M. (1996). Transcription factors in early development of the central nervous system. Neuropath Appl Neuro *22*, 469-481.

Sugimori, M., Nagao, M., Bertrand, N., Parras, C.M., Guillemot, F., and Nakafuku, M. (2007). Combinatorial actions of patterning and HLH transcription factors in the spatiotemporal control of neurogenesis and gliogenesis in the developing spinal cord. Development *134*, 1617-1629.

Sun, T., Dong, H.L., Wu, L.Z., Kane, M., Rowitch, D.H., and Stiles, C.D. (2003). Cross-repressive interaction of the Olig2 and Nkx2.2 transcription factors in developing neural tube associated with formation of a specific physical complex. J Neurosci *23*, 9547-9556.

Thomson, M., Liu, Siyuan J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. Cell *145*, 875-889.

Trompouki, E., Bowman, T.V., Lawton, L.N., Fan, Z.P., Wu, D.C., DiBiase, A., Martin, C.S., Cech, J.N., Sessa, A.K., Leblanc, J.L.*, et al.* (2011). Lineage Regulators Direct BMP and Wnt Pathways to Cell-Specific Programs during Differentiation and Regeneration. Cell *147*, 577-589.

Vokes, S.A., Ji, H.K., McCuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh, W.J.R., Davidson, E.H., Wong, W.H., and McMahon, A.P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development *134*, 1977-1989.

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385-397.

Zhou, Q., and Anderson, D.J. (2002). The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. Cell *109*, 61-73.

Zhou, Q., Choi, G., and Anderson, D.J. (2001). The bHLH transcription factor Olig2 promotes oligodendrocyte differentiation in collaboration with Nkx2.2. Neuron *31*, 791-807.

Zhou, Q., Wang, S.L., and Anderson, D.J. (2000). Identification of a novel family of oligodendrocyte lineage-specific basic helix-loop-helix transcription factors. Neuron *25*, 331-343.

Chapter 4

Conclusions and Future Directions

Research Summary

In my thesis work, I utilized ESC-models of pluripotency and neural development to investigate genome-scale gene regulatory programming of biological processes. In the first study, I explored opposing actions of the Wnt pathway in maintenance of pluripotency and commitment of ESCs to early embryonic fates. In the second study, I used Shh-mediated patterning of neural progenitors to examine neural progenitor patterning, and neural progenitor programming. My focus was directed to a large extent around bioinformatics-based hypothesis building and hypothesis testing utilizing large collections of genomic datasets generated within the McMahon laboratory, or available from published work.

Wnt and pluripotency

We identified genomic targets of the Wnt signaling pathway co-activator β-catenin: the first such genome scale study of a Wnt activator program in ESC systems. The genome-wide characterization of β-catenin transcriptional targets allow us to discover a strong association between β-catenin and core pluripotency factors, represented by Nanog, Oct4, Sox2, and Tcf3. We were also able to mechanistically illustrate β-catenin involvement in maintaining pluripotency via active input onto Oct/Sox chromatin binding sites via a protein complex with Oct4 and Tcf3. This additional mode of pluripotency maintenance interchanges dynamically with the Oct4/Sox2-mediated known mode of pluripotency maintenance, in the presence of active β-catenin. Meanwhile, we observed the inhibition of differentiation gene activation via ERK/Ets pathway suppression, without altering DNA-binding of β-catenin/Tcf around target genes of differentiation programs. These studies provide a molecular explanation of how two inhibitory small molecules may maintain ESC pluripotency under defined serum-free conditions. The two

factors are the GSK3 inhibitor CHIR99021 (CHIR) and ERK pathway inhibitor PD0325901 (PD03).

We propose that in this two inhibitors (abbrev: 2i) culture system, where β-catenin-directed canonical Wnt pathway activation is triggered by CHIR, and Ets family TF action is inhibited by PD03, β-catenin-dependent differentiation targets fail to be activated because of simultaneous requirement for Ets-factor engagement at *cis*-regulatory elements. Fgf-signaling is known to regulate ES differentiation, FGFs are secreted by ESCs and FGFs signal through Ets family members by the MEK pathway.

Some preliminary experimental evidence supports this model. In a genetically modified mESC line in which the Fgf-dependent Ets-factor Etv4 is fused to an engrailed domain (Etv4EnR) (Figure 4.1A), alkaline phosphatase positive cultures indicative of pluripotency are maintained in the absence of PD03, in contrast to wild type mESCs (Figure 4.1B-C). Since Etv4EnR is expected to bind to the same targets as Etv4/5, both expressed in ES cells, but to repress rather than activate targets due to the repressive engrailed domain, the result are consistent with PD03 inhibition of Ets factor-mediated transcriptional activation of differentiation target genes.

In addition, we have also shown that for genes associated with differentiation of ESCs to mesoderm and trophectodermal lineages, T and Cdx2, respectively, Etv4EnR expression is sufficient to inhibit T and Cdx2 up-regulation in PD03 withdrawn 2i culture (CHIR+/PD03-). Expression of T, Cdx2, and Cdx4 increases upon the removal of PD03. Induced expression of Etv4EnR significantly down-regulated the expressions of T and Cdx2 in 2i-adapted mESCs cultured PD03 withdrawn 2i culture (CHIR+/PD03-) (Figure 4.1D, lower panel). In contrast, the

Figure 4.1

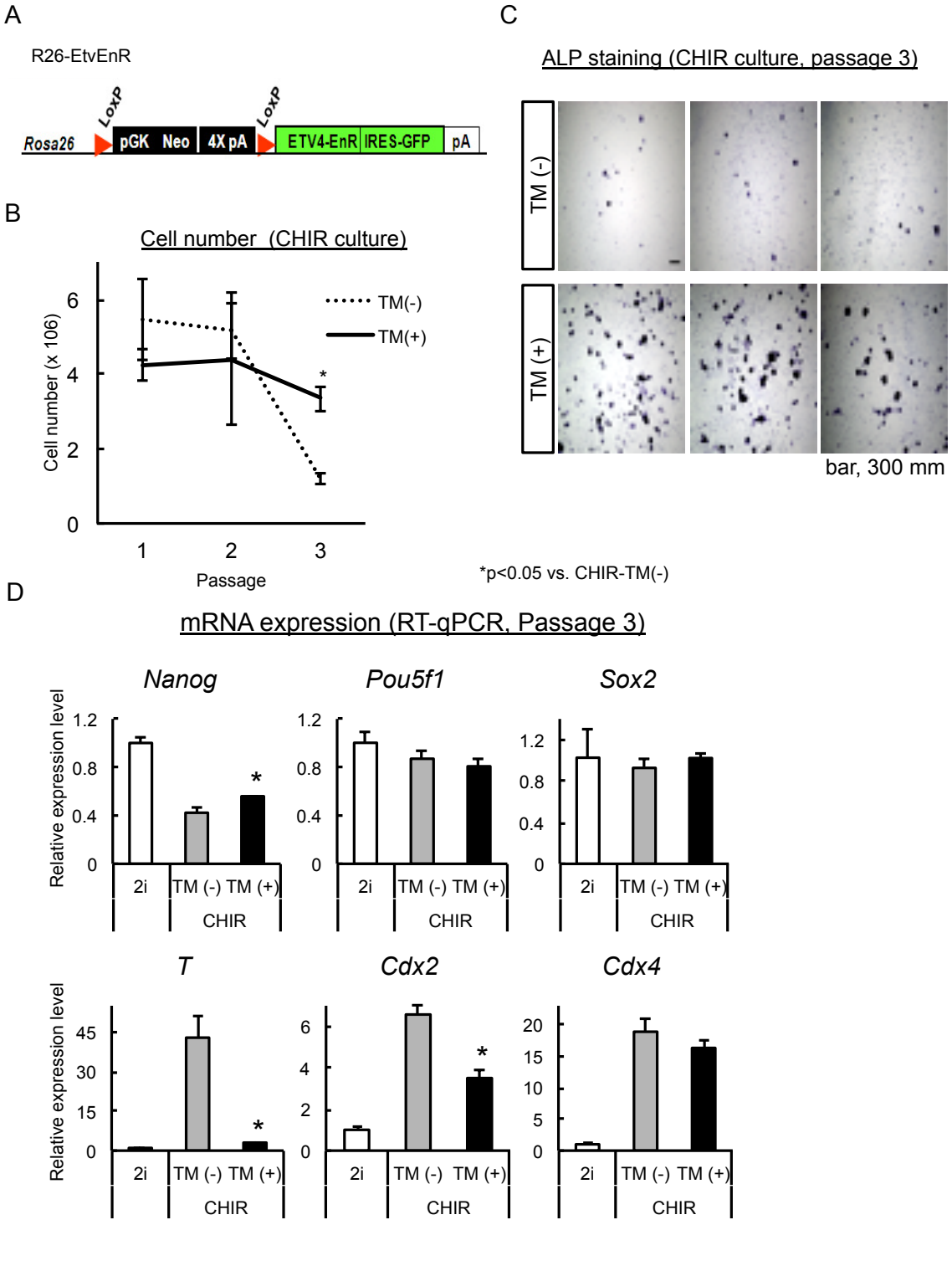

A
R26-EtvEnR
Rosa26
LoxP
pGK
Neo
4X pA
LoxP
ETV4-EnR
IRES-GFP
pA
B
Cell number (CHIR culture)
Cell number (x 106)
TM(-)
TM(+)
Passage
C
ALP staining (CHIR culture, passage 3)
TM (-)
TM (+)
bar, 300 mm
*p<0.05 vs. CHIR-TM(-)
D
mRNA expression (RT-qPCR, Passage 3)
Nanog
Pou5f1
Sox2
T
Cdx2
Cdx4
Relative expression level
2i
TM (-)
TM (+)
CHIR

Figure 4.1 (Continued)
Preliminary results of PD03-independent ESC culture for Etv4-EnR ES cell line (courtesy of Dr. Ohba, unpublished).

(A) Schematic of Etv4-EnR knock-in at Rosa26 locus flanked by loxP site (Mao et al., 2009). The Etv4-EnR locus will be transcribed when 4-OH-Tamoxifen is added to the cell culture.

(B) Comparison of survivals of parental line and 4-OH-TM-treated Etv4EnR-expressing cell line in CHIR-only 2i culture system (PD03 withdrawn). Cell number at passage-3 was significantly lower in parental cells than in 4OH-TM-treated (Etv4EnR-expressing) ones.

(C) Alkaline phosphatase staining showing the pluripotency of cell culture from parental line and 4-OH-TM-treated cell line at passage-3. Significant AP+ colonies were observed for Etv4EnR-expressing cells even after withdrawing PD03 for 3 passages.

(D) Expression change of representative pluripotency markers (top row) and differentiation genes (bottom row) in parental and 4OH-TM-treated cells before and after PD03 withdrawal. In contrast, removal of PD03 did not affect Pou5f1 and Sox2 expressions compared to 2i, although Nanog expression decreased dramatically upon the removal of PD03.

expression levels of pluripotency genes such as Oct4 and Sox2 were not altered on Etv4EnR expression, consistent with our prediction that only differentiation genes are subjected to inhibition by PD03 in 2i culture. Nanog expression is decreased after PD03 withdrawal, consistent with published work of a more direct link between PD03-activity and Nanog induction (Silva et al., 2009).

Future experimental tests of our model would use ChIP-qPCR against Etv4EnR and β-catenin examining *cis*-elements near differentiation-related genes. We would predict that Etv4 and β-catenin would co-localize at these *cis*-elements and cooperatively up-regulate differentiation gene expression. The next step will be to functionally test regulation of key differentiation genes by Etv4 and β-catenin. This could be measured in luciferase assays, where luciferase is under the regulation of an early differentiation target, such as T. We expect to see luciferase signals when both Etv4 and β-catenin are co-transfected, which will be abolished when the Lef/Tcf motif is mutated, or Etv4EnR instead is present.

Though there are concerns with individual experimental approaches that have been used to date: pharmacological inhibitor of GSK3, Wnt3a conditional medium treatment, and β-catenin or TCF3 modulation; there is good agreement amongst the different systems. Over-expression of β-catenin stimulates self-renewal of mESCs in a similar way to Wnt3a supplementation of medium (Ogawa et al., 2006; Takao et al., 2007). Moreover, GSK3 inhibition by knockout, exogenous expression of β-catenin, as well as Wnt3a treatment all resulted in similar activation of endogenous target genes or transfected reporter plasmids (Cole et al., 2008; Doble et al., 2007; Kelly et al., 2011; Kielman et al., 2002; Sato et al., 2004; Yi et al., 2008), pointing to β-catenin action in each experimental condition.

Some controversy exists on how β-catenin exerts its pro-self-renewal effect in mESCs. One potential mechanism is Tcf-independent β-catenin function via β-catenin-Oct4 complexes (Kelly et al., 2011). In other words, there is a separation of Tcf-dependent β-catenin transcriptional activities from Tcf-independent β-catenin targets. Supporting evidence for Tcf-independent mechanism also came from evidence that mESCs over-expressing a β-catenin-ΔC mutant, incapable of co-activator activity but capable of Tcf interaction, can stimulate self-renewal by inhibiting Tcf3 repressor activity (Wray et al., 2011; Yi et al., 2011). However, it was also proposed that Oct4 was not capable to bring β-catenin to chromatin, and Tcf factors might be required for β-catenin chromatin binding. Our model is consistent with the view that β-catenin is closed associated with core pluripotency factor Oct4. We further proposed that Tcf factor(s) are required for β-catenin chromatin binding via the Sox2 recognition component of a Oct4/Sox2 composite DNA-binding motif. In this, the Sox2 component of the DNA recognition site accounts for CHIR-mediated active input. The Oct4/β-catenin/Tcf-dependent pluripotency gene regulation is a separate, though closely intertwined, mechanism compared with the traditional Oct4/Sox2/Nanog core pluripotency network.

Shh and neural patterning

In early neural decision-making we proposed that the SoxB1 family member, Sox2, acts as a neuronal lineage-priming factor essential for a neural-specific Shh response. In this, Gli transcriptional effectors are proposed to regulate a number of VNT cell type determinants (Class II target genes). In the first step towards testing this model, we systematically recovered and validated CRMs for all Class II genes, and demonstrated Sox2 binding at these sites in neural progenitors. Further, Gli action was dependent on Sox2 binding for at least some of these genes.

Future experimental work will involve temporal chromatin binding profiles of Sox2 and Gli1 via ChIP-qPCR or ChIP-seq. The prediction is that if indeed Sox2 priming is required prior to Gli1 binding, we expect to see a temporal order of Sox2 binding precedence of Gli1.

We also observed extensive overlap of *cis*-regulatory elements utilized by ventral determinants activated through Gli-dependent processes: Nkx2.2, Nkx6.1, and Olig2. The system level characterization of common *cis*-elements deepens our understanding of cross-repression-mediated ventral progenitor cell layer patterning. Furthermore, by *in silico* characterization of bHLH family chromatin binding motif, E-box motif, we proposed a model of cell fate choice from neuronal progenitor cells to motor neuron (MN) or oligodendrocyte progenitor (OLP) cell fate difference via DNA binding sequence recognition site specificity.

More specifically, the model proposes Olig2 heterodimers promotes OLP cell fate by preferentially binding to CATCTG sequences activating target gene transcription in glial cells whereas Olig2::Olig2 homodimers have a preference for CATATG inhibiting gene activity in MN progenitors. Future experimental work to test this model will involve, first of all, the identification of Olig2 heterodimer species, and Ascl1 as well as NeuroD1/NeuroD4/Neurog2 are strong candidates for OLP and MN lineages, respectively. Secondly, the validation of site specificity by the proposed Olig2 heterodimers and the identification of functional Olig2 dimer species. This will be performed utilizing protein binding microarray technology for *in vitro* biochemical validation and ChIP-seq against tethered dimers by the *in vitro* neural embryoid body differentiation culture system. Thirdly, functional validation will be performed by ectopic expression of tethered dimers using chick electroporation. Fourthly, a quick method to validate the activating versus repressive roles of Olig2 dimers can be tested by luciferase assay.

While the general model may be correct, the model is likely to oversimplify these regulatory processes. For example, although Ascl1 expression correlates well with the generation of OLPs generation, and there is substantial decrease of OLPs in Ascl1-/- mutants, there is a recovery of OLPs in later stages. This suggests other bHLH factors may compensate for loss of Ascl1, such as Ascl3 and Ascl5 (Sugimori et al., 2008). Further, the generation of MN and OLP are under combinatorial actions of patterning factors including Nkx2.2, Olig2, and Pax6, and proneural bHLH factors, like Ascl1 and Neurogenin2. While we focus mostly on Olig2 and OLPs generated from within the pMN pool, OLPs can be generated from other progenitor cells (Sugimori et al., 2008).

An interesting question is whether multi-potent progenitors first commit to neuron- and glia-restricted progenitors, which later differentiate into astrocytes and oligodendrocytes, or alternatively, neuron-oligodendrocyte progenitors dominate, together with separate astrocyte progenitors (Anderson et al., 2002; Gabay et al., 2003; Lu et al., 2002; Pringle et al., 2003). Some evidence argues for different domains in the VNT that generate different cell types in a distinct progression, with a pool of undifferentiated progenitors persisting throughout the course of neuro/gliogenesis (Sugimori et al., 2007).

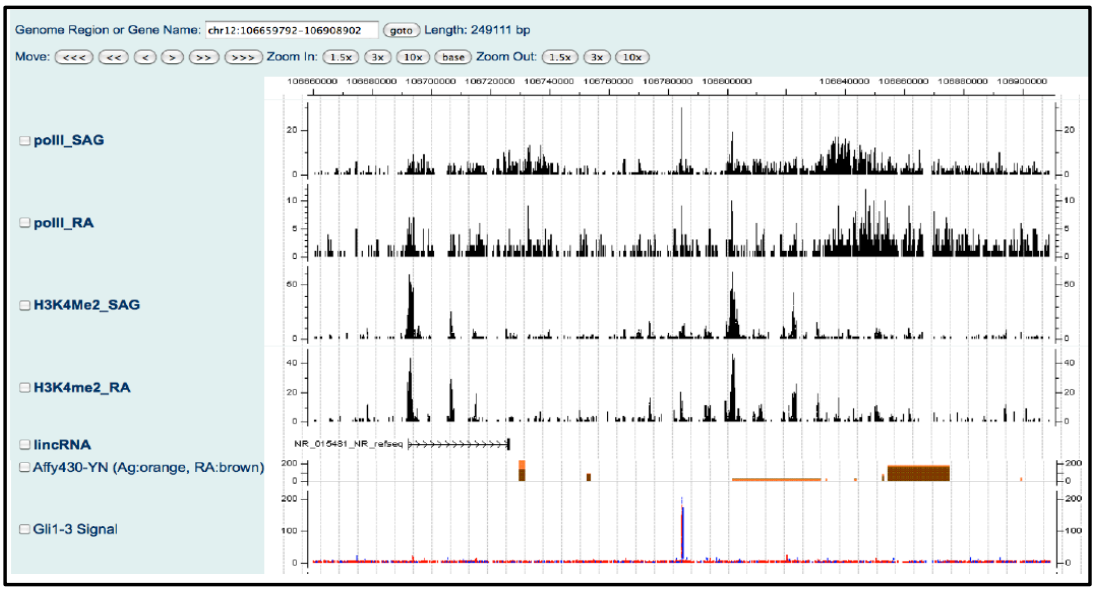
In the future, with the time course RNA-seq data in hand, we could extend our investigation of Shh-responsive genes to the potential actions of regulatory RNAs in Shh-mediated neural patterning, such as long inter-genic non-coding (linc) RNA and microRNA (miRNA) transcripts. Non-coding miRNAs and lincRNAs play important regulatory roles in gene networks during development (Guttman et al., 2009; Guttman et al., 2011). While many functions have been ascribed to miRNAs, and miRNA gene organization is well described, lincRNAs are less well understood. So far the mammalian genome is predicted to encode ~3500

lincRNAs. lincRNAs have roles in a broad spectrum of activities, such as chromatin structure modification, transcriptional regulation by recruiting TFs, post-transcriptional RNA processing, and translational controls (Qureshi et al., 2010). Guttman et al. (2011) performed a systematic functional study on all known ESC-expressed lincRNAs by knockdown experiments, and found dozens of lincRNAs that are essential for ESCs to remain pluripotent or to repress lineage-specific gene expressions (Guttman et al., 2011). In the neural system, lincRNAs have been found to be involved in mediating cognitive and behavioral processes (Qureshi and Mehler, 2011). A potential intersection between Shh-Gli regulation and miRNA and lincRNA encoding targets has not been addressed and would complete our general analysis of regulatory strategies.

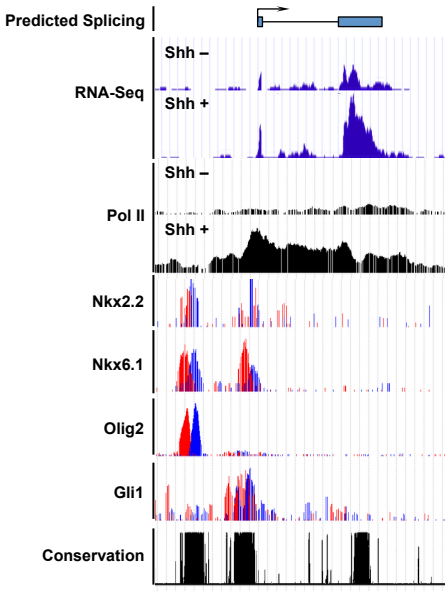
From preliminary analysis, my colleagues have shown that regulatory RNAs, miRNAs and lincRNAs, also intersect with our primary data; 14 GBRs lie next to a gene encoding a miRNA, while 142 have a lincRNA gene as a nearest 5' or 3' neighbor (Figure 4.2A). Thus, both classes of regulatory RNAs are potential downstream regulators of Shh action. However, determining whether this may be the case requires a specific analysis of regulatory RNAs in our model. As an example of the approach in identifying Shh-responsive lincRNA (Figure 4.2B), this lincRNA has very strong signal in the Shh-positive sample, and has strong ventral TF binding in the vicinity. By in situ hybridization, this transcript shows nice expression in the progenitor regions of the VNT (Figure 4.2C). In the future, functional assays, including ectopic expression and knock down of regulatory RNAs in mESC system or conditional mutation in mouse models can be utilized to assist in the discovery and experimental analysis of Shh-responsive regulatory RNA in early embryonic NT patterning.

Figure 4.2

A



B



C

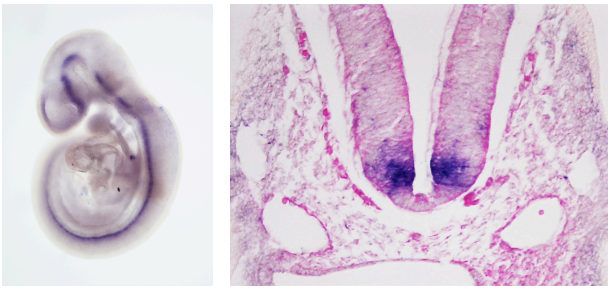


Figure 4.2 (Continued)
Examples of Shh-responsive lincRNA from RNA-seq data (Courtesy of Dr. Nishi, Dr. Xu, Rinn Lab)
(A) CisGenome browser view of Gli1-FLAG binding associated with a predicted linc RNA encoding gene following SAG treatment of neuralized EBs . A strong GBS (lowest track) is located ~ 50 kb downstream of NR_015481, a predicted lincRNA-encoding gene. The GBR shows an elevated me2 and polII signature and the predicted lincRNA gene displays a strong H3K4me2 signature at the promoter region. An Affymetrix probe-set just 3' to lincRNA gene structure prediction shows a two-fold enhanced expression on SAG-mediated Shh pathway activation (orange + SAG and brown – SAG). LincRNA dataset was provided by Dr. J. Rinn Laboratory.
(B) CisGenome browser example of a novel predicted Shh-responsive lincRNA. Chromatin binding signals from Gli1 and ventral progenitor determinants are also shown to demonstrate potential regulation of lincRNA transcription by Shh network genes.
(C) Whole mount and cross-sectional in situ hybridization of the transcript in (B) showing strong expression in the progenitor domains of VNT, fulfilling the spatial requirement of Shh-responsiveness.

Future challenges and directions for studying gene regulation on system level

All sequencing technologies have some biases and artifacts. For example, there is bias towards GC-rich content in fragment selection. Sequencing errors are almost always present, although advance in technology has largely reduced this factor. How many reads need to be sequenced in order to have a reasonable coverage of the genome is always a heated debate in the field (Kharchenko et al., 2008), which is especially crucial for histone modification ChIP-seq that requires a large number of short reads sequencing. Finally, bioinformatics-wise, data management, such as storage, sharing, and processing is a unique challenge for next generation sequencing (NGS) data, and luckily, large public database, such as Gene Expression Omnibus (Edgar et al., 2002), is gradually building up and standardizing data from various labs and consortiums.

For ChIP-seq, relatively large amounts of starting material are needed, which can be prohibitive for some experiments. Further, antibody quality determines the feasibility and specificity of the experiment. Although epitope tags can always be one solution, the availability of epitope-tagged protein of interest proposes another major challenge. One inherent disadvantage is that the profiles identified represent an average binding status for a given factor in the whole cell population. While this is not a major problem for TF ChIP-seq, for histone modification ChIP-seq and TFs with broad expression domains, conclusions on co-binding of multiple factors need to be drawn with caution. In addition, many TF binding *cis*-elements are located in intergenic regions that are more than 10kb from transcription start site (TSS) of target genes. For example, only 12% of GBRs occurred within 10 kb of a TSS, suggesting that long-range interactions are a common feature of the Gli1 regulatory program. Similarly, only 5.3% of

the Nkx2.2+Nkx6.1+Olig2+ *cis*-regulatory elements are within 10 kb of a TSS. Associating a binding region with its target genes has been a tough problem for all peak-calling algorithms.

Traditionally, binding sites are assigned to the nearest genes, but research has shown that such simplified association introduces a strong bias for genes in large gene dessert. Moreover, several enhancer elements located >0.5 Mb from target gene TSS have been reported. For example, a Gli1 binding site located 540 kb upstream of Nkx6.1 TSS has been validated by transgenic assay. One mechanism for bringing these long-range enhancers to specific promoters have been assigned to insulators, which can bring together sequences located far apart in the linear genome (Yang and Corces, 2012).

Complications of bioinformatics analyses can also come from 3-dimentional (3D) chromatin structures variations, such as looping. Recent advances in Chromosomal Conformation Capture (3C)-based techniques, such as 4C, Hi-C, and 5C, have unveiled the 3D architecture of the genome and the compartmentalization of the nucleus, as well as the phenomenon that enhancers far apart and even on different chromosomes can be folded into proximity (Lieberman-Aiden et al., 2009). Therefore various efforts have been devoted to establish a rule for assigning *cis*-regulatory elements to the real target promoters. For example, GREAT algorithm associates *cis*-elements with target genes by defining a 'regulatory domain' for each gene in the genome, ranging from 5 kb upstream and 1 kb downstream of gene TSS, and an extension up to the basal regulatory domain of the nearest upstream and downstream genes within 1 Mb as default (McLean et al., 2010). When computing for gene ontology term enrichments, a binomial test is further performed to adjust biases introduced by variability in gene regulatory domain sizes. In another study, Tang et al. (2011) empirically defined the 'regulatory potential' of a TF for a given gene by taking into account of the number of TF

binding peaks, weighted by the distance from each site to the TSS, within a certain distance (default, 100 kb), without considering actual binding peak strength (Tang et al., 2011). Distinguishing functional binding from non-functional binding simply by chance, or affinity, is labor-intensive.

For RNA-seq, sensitivity of transcriptome and variations among samples is a significant issue biological replicates are important. Current sample processing protocols have several variation, such as whether to fragment RNA (RNA hydrolysis or nebulization) or cDNA (DNase I treatment or sonication), whether to enrich mRNA by poly(A)+ or by ribosomal depletion. Bioinformatics-wise, the complexity of eukaryote transcriptomes and the variation in RNA splicing makes the subsequent data processing challenging, such as the assembly of the short reads into contigs and the alignment of contigs with genomic sequences. As a final stage, the validation of *in silico* discovered transcript, especially low abundance ones, is also a challenge in reducing false positive hits.

As powerful as NGS technique is, the complexity of genomics data poses considerable challenges to biologists who are not familiar with system level type of data. This points to the increasing importance in cross-disciplinary communication and collaboration among biologists with statistician, mathematician, and computer scientists. In addition, while the plethora of bioinformatics algorithm, pipelines, and workflows provide users with ample selections for data analysis, there is no golden standard as to which method is the best for which data set. So the reproducibility of bioinformatics analysis can be a severe hurdle for communications in the field. Thus the open source sharing platform for NGS data as well as the standardization of NGS data analysis platform are at least as important as generating data in the long run.

References

Anderson, D.J., Choi, G., and Zhou, Q. (2002). Olig genes and the genetic logic of CNS neural cell fate determination. Clin Neurosci Res *2*, 17-28.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Gene Dev *22*, 746-755.

Doble, B.W., Patel, S., Wood, G.A., Kockeritz, L.K., and Woodgett, J.R. (2007). Functional Redundancy of GSK-3α and GSK-3β in Wnt/β-Catenin Signaling Shown by Using an Allelic Series of Embryonic Stem Cell Lines. Dev Cell *12*, 957-971.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res *30*, 207-210.

Gabay, L., Lowell, S., Rubin, L.L., and Anderson, D.J. (2003). Deregulation of dorsoventral patterning by FGF confers trilineage differentiation capacity on CNS stem cells in vitro. Neuron *40*, 485-499.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P.*, et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L.*, et al.* (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295-U260.

Kelly, K.F., Ng, D.Y., Jayakumaran, G., Wood, G.A., Koide, H., and Doble, B.W. (2011). β-Catenin Enhances Oct-4 Activity and Reinforces Pluripotency through a TCF-Independent Mechanism. Cell Stem Cell *8*, 214-227.

Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat Biotechnol *26*, 1351-1359.

Kielman, M.F., Rindapää, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., Taketo, M.M., Roberts, S., Smits, R., and Fodde, R. (2002). Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. Nat Genet *32*, 594-605.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science *326*, 289-293.

Lu, Q.R., Sun, T., Zhu, Z.M., Ma, N., Garcia, M., Stiles, C.D., and Rowitch, D.H. (2002). Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection. Cell *109*, 75-86.

Mao, J., McGlinn, E., Huang, P., Tabin, C.J., and Mcmahon, A.P. (2009). Fgf-dependent Etv4/5 activity is required for posterior restriction of Sonic Hedgehog and promoting outgrowth of the vertebrate limb. Dev Cell *16*, 600-606.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-U155.

Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D., and Niwa, H. (2006). Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. Biochem Biophys Res Commun *343*, 159-166.

Pringle, N.P., Yu, W.-P., Howell, M., Colvin, J.S., Ornitz, D.M., and Richardson, W.D. (2003). Fgfr3 expression by astrocytes and their precursors: evidence that astrocytes and oligodendrocytes originate in distinct neuroepithelial domains. Development *130*, 93-102.

Qureshi, I.A., Mattick, J.S., and Mehler, M.F. (2010). Long non-coding RNAs in nervous system function and disease. Brain Research *1338*, 20-35.

Qureshi, I.A., and Mehler, M.F. (2011). Non-coding RNA networks underlying cognitive disorders across the lifespan. Trends in Molecular Medicine *17*, 337-346.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med *10*, 55-63.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., Oosten, A.L.v., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. Cell *138*, 722-737.

Sugimori, M., Nagao, M., Bertrand, N., Parras, C.M., Guillemot, F., and Nakafuku, M. (2007). Combinatorial actions of patterning and HLH transcription factors in the spatiotemporal control of neurogenesis and gliogenesis in the developing spinal cord. Development *134*, 1617-1629.

Sugimori, M., Nagao, M., Parras, C.M., Nakatani, H., Lebel, M., Guillemot, F., and Nakafuku, M. (2008). Ascl1 is required for oligodendrocyte development in the spinal cord. Development *135*, 1271-1281.

Takao, Y., Yokota, T., and Koide, H. (2007). Beta-catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. Biochem Biophys Res Commun *353*, 699-705.

Tang, Q.Z., Chen, Y.W., Meyer, C., Geistlinger, T., Lupien, M., Wang, Q., Liu, T., Zhang, Y., Brown, M., and Liu, X.S. (2011). A Comprehensive View of Nuclear Receptor Cancer Cistromes. Cancer Res *71*, 6940-6947.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol *13*, 838-U246.

Yang, J.P., and Corces, V.G. (2012). Insulators, long-range interaction, and genome function. Curr Opin Genet Dev *22*, 86-92.

Yi, F., Pereira, L., Hoffman, J.A., Shy, B.R., Yuen, C.M., Liu, D.R., and Merrill, B.J. (2011). Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nat Cell Biol *13*, 762-770.

Yi, F., Pereira, L., and Merrill, B.J. (2008). Tcf3 Functions as a Steady-State Limiter of Transcriptional Programs of Mouse Embryonic Stem Cell Self-Renewal. Stem Cells *26*, 1951-1960.